AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXXIV

1952

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

Display
I

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXXIV, Number 1
JANUARY, 1952

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

7

CONTENTS

The AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications should be sent to Professor AUREL WINTNER at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE 18, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

ON THE PICARD VARIETIES ATTACHED TO ALGEBRAIC VARIETIES.*

By Jun-ichi Igusa.

In spite of the numerous classical investigations the theory of Picard varieties is not so complete as the special case of Jacobian varieties. In order to illustrate this circumstance, we should like to mention the problem of the relation between the abelian variety attached to the period matrix of the Picard integrals of the first kind on the given algebraic variety and its Picard variety. We could not find any treatment of such a fundamental problem in the literature.[1]

Now in modern algebraic geometry we have the powerful tool of "harmonic integrals," which has been invented by de Rham and Hodge and fully investigated by Kodaira and de Rham recently, together with the rigorous foundation of algebraic geometry due to van der Waerden, Chow, Zariski and Weil. On the base of these achievements we shall develop a theory of Picard varieties which includes the complete solution of the above problem.

In § I, we shall prove (for the sake of completeness), that continuous equivalence in the sense of Severi implies homological equivalence over integers. In § II, we shall define the Albanese variety A attached to a given variety V and also a skew-symmetric integral matrix E. In § III, we shall define a continuous family of divisors on V, which is "parametrized" by A, after Poincaré's method. The Picard variety $\mathfrak{P}$ attached to V is then defined by a certain "involution" of A, which can be described by the above matrix E. Our results include Weil's duality between abelian varieties. This paper contains two Appendices; in the second we shall sketch a correspondence theory between algebraic varieties.

While the outline of this paper was made in Tokyo in August 1949, I have received in writing it down at Kyoto constant encouragement and kind advice from Prof. Akizuki and valuable suggestions from Prof. Weil, to whom I wish to express my deepest appreciation.

---

* Received July 12, 1950; revised June 13, 1951.

[1] Indeed these two abelian varieties have been confused by some authors perhaps from the fact that they coincide if the given variety is a curve. On the other hand I heard from Weil by his kind letter of October 19, 1949 that he had also remarked the distinction between the abelian variety and its Picard variety; see § III, 10 of this paper.

§ I. Continuous Families and Continuous Equivalences on a Projective Model V.

1. Let L^n be a projective space in the algebraic geometry with the universal domain K of all complex numbers.[2] Let H^{n-r} be a linear variety in L^n, defined by a set of equations

$$\sum_{j=0}^{n} u_{ij}X_j = 0 \qquad (1 \leqq i \leqq r),$$

where $(u) = (u_{10}, u_{11}, \cdots, u_{rn})$ is a set of $r(n+1)$ independent variables over a field k. Such a linear variety was systematically used for the first time by van der Waerden; we shall call it a *generic linear variety over* k.

Now let V^d be any compact complex analytic variety of (complex) dimension d in L^n; by a recent result of Chow,[3] V is then an algebraic variety. We shall assume that V has no multiple point. Although we shall consider the case $d \geqq 2$, our results hold trivially in the case $d = 1$. In the following we shall assume that every field contains the smallest field of definition of V.

We shall start from the following lemma, which is slightly different from a similar lemma of Zariski.[4]

LEMMA. *Let L be a regular extension of a field K, and let ξ, η be two independent variables in L over K. If u is a variable over L, then $L(u)$ is regular over the field $K(u)(\xi + u\eta)$.*

Proof. As is readily seen, it is sufficient to consider the case where L is an algebraic extension over $K(\xi, \eta)$. Let u_i $(1 \leqq i \leqq N)$ be a set of N independent variables over L, and let K_i be the algebraic closure of $K(u_1, \cdots, u_N)(\xi + u_i\eta)$ in $L(u_1, \cdots, u_N)$. Then, since

$$[L(u_1, \cdots, u_N) : K(u_1, \cdots, u_N)(\xi, \eta)] = [L : K(\xi, \eta)],$$

if we take N sufficiently large, there must exist at least one pair (i, j) $(i \neq j)$ such that $K_i(\xi + u_j\eta) = K_j(\xi + u_i\eta)$. On the other hand since K_j is regular over $K(u_1, \cdots, u_N)$, $K_j(\xi + u_i\eta)$ is regular over $K(u_1, \cdots, u_N)(\xi + u_i\eta)$.

---

[2] We shall use the results and terminology of Weil's book: *Foundations of algebraic geometry*, American Mathematical Society Colloquium Publication no. 29 (1946). We shall cite this book as (F).

[3] W. L. Chow, "On compact complex analytic varieties," this JOURNAL, Vol. 71 (1949); theorem 5. We shall cite this paper as (C).

[4] O. Zariski, "Pencils on an algebraic variety and a new proof of a theorem of Bertini," *Transactions of the American Mathematical Society*, vol. 50 (1941); lemma 5.

Therefore $K_i = K(u_1, \cdots, u_N)(\xi + u_i\eta)$, and $L(u_1, \cdots, u_N)$ is regular over $K(u_1, \cdots, u_N)(\xi + u_i\eta)$; hence $L(u_i)$ is regular over $K(u_i)(\xi + u_i\eta)$.

Let $(x) = (x_0, x_1, \cdots, x_n)$ be a representative of a generic point of V over a field k, and say $x_0 = 1$; let $(u_1, \cdots, u_n)$ be a set of independent variables over $k(x)$, and put $u_0 = -\sum_{i=1}^{n} u_i x_i$. Then the equation $\sum_{j=0}^{n} u_i X_i = 0$ defines a generic linear variety H^{n-1} in L^n over k. Concerning the section of V by such an H, we have the following basic result.

PROPOSITION 1. *The intersection-product* $\mathsf{V} \cdot \mathsf{H}$ *is an (absolutely irreducible) variety without multiple point.*

Proof. Let x_{i_0} be a variable among the x_i $(1 \leqq i \leqq n)$ over k, and put

$$\xi = \sum_{0 \neq i \neq i_0} u_i x_i, \qquad \eta = x_{i_0};$$

then ξ, η are independent variables over $k(u_1, \cdots, u_{i_0-1}, u_{i_0+1}, \cdots, u_n)$, hence by the above lemma, $k(u_1, \cdots, u_n, x)$ is regular over $k(u_0, u_1, \cdots, u_n)$. Therefore the point of V with the representative (x) has a locus W over $k(u_0, u_1, \cdots, u_n)$ such that $\mathsf{V} \cdot \mathsf{H} = i(\mathsf{V} \cdot \mathsf{H}, \mathsf{W})\mathsf{W}$. On the other hand, H is transversal to V along W, hence by the criterion of multiplicity 1, we have $i(\mathsf{V} \cdot \mathsf{H}, \mathsf{W}) = 1$. Moreover by the "theorem of Bertini" on the variable singular points, W has no multiple point.

By repeated use of this proposition we conclude that the intersection-product W^{d-r} $(0 \leqq r < d)$ of V^d and a generic linear variety H^{n-r} over a field k is a subvariety of V without multiple points; we shall call it a *generic* $(d-r)$*-section of* V^d *over* k.

2. Now let Y be a positive r-cycle $(0 \leqq r \leqq d)$ on V and let M be the coefficients of the "associated form" of Y. With Chow we shall consider M as a set of homogeneous coordinates of a point in a projective space; we shall call it the *Chow point* of Y. We shall also identify M with the point itself. By an elementary property of the Chow point,[5] if M' is a specialization of M over a field k, there exists a rational r-cycle Y' with the Chow point M' over $k(\mathsf{M}')$. It is natural to call Y' a *specialization of* Y *over* k. Furthermore let Y be an r-cycle on V and let $\mathsf{Y} = \mathsf{Y}_1 - \mathsf{Y}_2$ be the reduced expression of Y as a difference of positive cycles Y_1 and Y_2 then if Y'_1 and Y'_2 are the specializations of Y_1 and Y_2 over k, we say that $\mathsf{Y}' = \mathsf{Y}'_1 - \mathsf{Y}'_2$ is a specializa-

---

[5] For a systematic treatment of the Chow point see v. d. Waerden's book: *Einführung in die algebraische Geometrie*, Berlin (1939), §§ 36-38.

tion of Y over k. On the other hand let K be an extension of k over which both Y and Y' are rational, and let H^{n-r} be a generic linear variety over K. It can be readily seen that Y' is a specialization of Y over k if and only if the 0-cycle $\mathsf{Y}' \cdot \mathsf{H}$ is a specialization of the 0-cycle $\mathsf{Y} \cdot \mathsf{H}$ over $k(u)$, where (u) denotes the set of coefficients of the equations for H. Since our specialization of cycles is based on the point specialization, the basic properties of the usual specializations are valid for this general specialization; we shall show

PROPOSITION 2. *Let* U *be an abstract variety defined over a field* k; *let* M *be a generic point of* U *over* k *and let* M' *be a simple point of* U. *Let* X *be a rational* $(\mathsf{U} \times \mathsf{V})$*-cycle over* k *such that the* V*-cycle*

$$\mathsf{X}(\mathsf{M}') = \mathrm{pr}_{\mathsf{V}}(\mathsf{X} \cdot (\mathsf{M}' \times \mathsf{V}))$$

is defined. Then the similar cycle $\mathsf{X}(\mathsf{M})$ *has a uniquely determined specialization* $\mathsf{X}(\mathsf{M}')$ *over* $\mathsf{M} \to \mathsf{M}'$ *with reference to* k.

Proof. Let Y^r be a specialization of $\mathsf{X}(\mathsf{M})$ over $\mathsf{M} \to \mathsf{M}'$ with reference to k, and let K be an extension of $k(\mathsf{M}, \mathsf{M}')$ over which Y is rational; let H^{n-r} be generic linear variety over K. Then $\mathsf{Y} \cdot \mathsf{H}$ is a specialization of $\mathsf{X}(\mathsf{M}) \cdot \mathsf{H}$ over $\mathsf{M} \to \mathsf{M}'$ with reference to $k(u)$. On the other hand let W be the generic $(d-r)$-section of V by H over k, then we have

$$\mathsf{X}(\mathsf{M}) \cdot \mathsf{H} = \mathrm{pr}_{\mathsf{W}}((\mathsf{X}_0 \cdot (\mathsf{M}' \times \mathsf{W}))_{\mathsf{U}\times\mathsf{W}}),$$

$$\mathsf{X}(\mathsf{M}') \cdot \mathsf{H} = \mathrm{pr}_{\mathsf{W}}((\mathsf{X}_0 \cdot (\mathsf{M} \times \mathsf{W}))_{\mathsf{U}\times\mathsf{W}}),$$

where $\mathsf{X}_0 = \mathsf{X} \cdot (\mathsf{U} \times \mathsf{W})$. Therefore by Th. 13 of (F), Chap. VII, § 6, the W-cycle $\mathsf{X}(\mathsf{M}) \cdot \mathsf{H}$ has the *uniquely determined* specialization $\mathsf{X}(\mathsf{M}') \cdot \mathsf{H}$ over $\mathsf{M} \to \mathsf{M}'$ with reference to $k(u)$, hence we must have $\mathsf{Y} \cdot \mathsf{H} = \mathsf{X}(\mathsf{M}') \cdot \mathsf{H}$. However since Y and $\mathsf{X}(\mathsf{M}')$ are rational over K, and since H is a generic linear variety over K, this implies $\mathsf{Y} = \mathsf{X}(\mathsf{M}')$, which completes our proof.

Now let U, X, k and M be the same as in Proposition 2. but U be in the following a complete variety; then the specialization of $\mathsf{X}(\mathsf{M})$ over k constitute what we call the *continuous family determined by* X *on* V, or shortly the *continuous family* X. It can be readily verified that this definition does not depend on the choice of k and M, when X is given. Moreover we can and shall assume that every component of X has the projection U on U. We say that two r-cycles Y and Y' on V are *continuously equivalent*, if there exists a series of r-cycles $\mathsf{Y} = \mathsf{Y}_1, \mathsf{Y}_2, \cdots, \mathsf{Y}_{N+1} = \mathsf{Y}'$, a series of complete non-singular curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_N$ and a series of $(\Gamma_i \times \mathsf{V})$-cycles X_i, such that

$$\mathsf{Y}_i = \mathrm{pr}_{\mathsf{V}}(\mathsf{X}_i \cdot (\mathsf{M}_i \times \mathsf{V})), \qquad \mathsf{Y}_{i+1} = \mathrm{pr}_{\mathsf{V}}(\mathsf{X} \cdot (\mathsf{M}'_i \times \mathsf{V})),$$

where M_i and M'_i are the points of Γ_i for $1 \leqq i \leqq N$. The continuous equivalence is thus an equivalence relation in the set of all r-cycles on V. Moreover since the points M_i and M'_i are homologous on Γ_i, the products $M_i \times V$ and $M'_i \times V$ are homologous on $\Gamma_i \times V$. Therefore the intersection products $X_i \cdot (M_i \times V)$ and $X_i \cdot (M'_i \times V)$ are homologous on $\Gamma_i \times V$, and their algebraic projections Y_i and Y_{i+1} on V are homologous on V. It follows, finally that

$$Y - Y' = \sum_{i=1}^{N} (Y_i - Y_{i+1})$$

is homologous to 0 on V. Here the homologies are taken over Z;[6] and all these follow from the topological intersection-theory.[7]

3. Now we shall connect our two notions by the following assertion.

THEOREM 1. *Any two cycles in the same continuous family on* V *are continuously equivalent.*

Proof. We shall use the same notations as before. Let Y be any specialization of $X(M)$ over k; we have only to show that they are continuously equivalent. Since U is a complete variety, there exists a point M' on U such that Y is a specialization of $X(M)$ over $M \to M'$ with reference to k. We shall first assume that U is a curve and let Γ be derived from U by normalization with reference to k; then Γ is a complete non-singular curve, which corresponds to U by a birational correspondence T over k. Let $M \times M^*$ be the generic point of T over k with the projection M on U; and let $M^{*\prime}$ be a specialization of M^* over the specialization $X(M) \to Y$, $M \to M'$ with reference to k. If we interchange the second and the third factors of the product $U \times \Gamma \times V \times V$, its subvariety $T \times \Delta$ corresponds biregularly to a birational correspondence Z between $U \times V$ and $\Gamma \times V$. Since Z is biregular along every component of X, the intersection-product $Z \cdot (X \times \Gamma \times V)$ is defined; we put $X^* = \mathrm{pr}_{\Gamma \times V}(Z \cdot (X \times \Gamma \times V))$. Then it can be readily verified that we have

$$X^*(M^*) = \mathrm{pr}_V (X^* \cdot (M^* \times V)) = X(M).$$

Moreover since every component of X^* has the projection Γ on Γ, the intersection-product $X^* \cdot (M^{*\prime} \times V)$ is defined. Therefore by the previous

---

[6] As in Bourbaki we shall denote by Z the ring of rational integers and by Q its field of quotients.

[7] Cf. S. Lefschetz, *Topology*, American Mathematical Society Colloquium Publication no. 12 (1930).

tion of Y over k. On the other hand let K be an extension of k over which both Y and Y' are rational, and let H^{n-r} be a generic linear variety over K. It can be readily seen that Y' is a specialization of Y over k if and only if the 0-cycle $\mathsf{Y}' \cdot \mathsf{H}$ is a specialization of the 0-cycle $\mathsf{Y} \cdot \mathsf{H}$ over $k(u)$, where (u) denotes the set of coefficients of the equations for H. Since our specialization of cycles is based on the point specialization, the basic properties of the usual specializations are valid for this general specialization; we shall show

PROPOSITION 2. *Let U be an abstract variety defined over a field k; let M be a generic point of U over k and let M' be a simple point of U. Let X be a rational $(\mathsf{U} \times \mathsf{V})$-cycle over k such that the V-cycle*

$$\mathsf{X}(\mathsf{M}') = \mathrm{pr}_{\mathsf{V}}(\mathsf{X} \cdot (\mathsf{M}' \times \mathsf{V}))$$

is defined. Then the similar cycle $\mathsf{X}(\mathsf{M})$ has a uniquely determined specialization $\mathsf{X}(\mathsf{M}')$ over $\mathsf{M} \rightarrow \mathsf{M}'$ with reference to k.

Proof. Let Y^r be a specialization of $\mathsf{X}(\mathsf{M})$ over $\mathsf{M} \rightarrow \mathsf{M}'$ with reference to k, and let K be an extension of $k(\mathsf{M}, \mathsf{M}')$ over which Y is rational; let H^{n-r} be generic linear variety over K. Then $\mathsf{Y} \cdot \mathsf{H}$ is a specialization of $\mathsf{X}(\mathsf{M}) \cdot \mathsf{H}$ over $\mathsf{M} \rightarrow \mathsf{M}'$ with reference to $k(u)$. On the other hand let W be the generic $(d-r)$-section of V by H over k, then we have

$$\mathsf{X}(\mathsf{M}) \cdot \mathsf{H} = \mathrm{pr}_{\mathsf{W}}((\mathsf{X}_0 \cdot (\mathsf{M}' \times \mathsf{W}))_{\mathsf{U} \times \mathsf{W}}),$$

$$\mathsf{X}(\mathsf{M}') \cdot \mathsf{H} = \mathrm{pr}_{\mathsf{W}}((\mathsf{X}_0 \cdot (\mathsf{M} \times \mathsf{W}))_{\mathsf{U} \times \mathsf{W}}),$$

where $\mathsf{X}_0 = \mathsf{X} \cdot (\mathsf{U} \times \mathsf{W})$. Therefore by Th. 13 of (F), Chap. VII, § 6, the W-cycle $\mathsf{X}(\mathsf{M}) \cdot \mathsf{H}$ has the *uniquely determined* specialization $\mathsf{X}(\mathsf{M}') \cdot \mathsf{H}$ over $\mathsf{M} \rightarrow \mathsf{M}'$ with reference to $k(u)$, hence we must have $\mathsf{Y} \cdot \mathsf{H} = \mathsf{X}(\mathsf{M}') \cdot \mathsf{H}$. However since Y and $\mathsf{X}(\mathsf{M}')$ are rational over K, and since H is a generic linear variety over K, this implies $\mathsf{Y} = \mathsf{X}(\mathsf{M}')$, which completes our proof.

Now let U, X, k and M be the same as in Proposition 2. but U be in the following a complete variety; then the specialization of $\mathsf{X}(\mathsf{M})$ over k constitute what we call the *continuous family determined by* X *on* V, or shortly the *continuous family* X. It can be readily verified that this definition does not depend on the choice of k and M, when X is given. Moreover we can and shall assume that every component of X has the projection U on U. We say that two r-cycles Y and Y' on V are *continuously equivalent*, if there exists a series of r-cycles $\mathsf{Y} = \mathsf{Y}_1, \mathsf{Y}_2, \cdots, \mathsf{Y}_{N+1} = \mathsf{Y}'$, a series of complete non-singular curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_N$ and a series of $(\Gamma_i \times \mathsf{V})$-cycles X_i, such that

$$\mathsf{Y}_i = \mathrm{pr}_{\mathsf{V}}(\mathsf{X}_i \cdot (\mathsf{M}_i \times \mathsf{V})), \qquad \mathsf{Y}_{i+1} = \mathrm{pr}_{\mathsf{V}}(\mathsf{X} \cdot (\mathsf{M}'_i \times \mathsf{V})),$$

where M_i and M'_i are the points of Γ_i for $1 \leqq i \leqq N$. The continuous equivalence is thus an equivalence relation in the set of all r-cycles on V. Moreover since the points M_i and M'_i are homologous on Γ_i, the products $M_i \times V$ and $M'_i \times V$ are homologous on $\Gamma_i \times V$. Therefore the intersection products $X_i \cdot (M_i \times V)$ and $X_i \cdot (M'_i \times V)$ are homologous on $\Gamma_i \times V$, and their algebraic projections Y_i and Y_{i+1} on V are homologous on V. It follows, finally that

$$Y - Y' = \sum_{i=1}^{N} (Y_i - Y_{i+1})$$

is homologous to 0 on V. Here the homologies are taken over Z;[6] and all these follow from the topological intersection-theory.[7]

3. Now we shall connect our two notions by the following assertion.

THEOREM 1. *Any two cycles in the same continuous family on* V *are continuously equivalent.*

Proof. We shall use the same notations as before. Let Y be any specialization of $X(M)$ over k; we have only to show that they are continuously equivalent. Since U is a complete variety, there exists a point M' on U such that Y is a specialization of $X(M)$ over $M \to M'$ with reference to k. We shall first assume that U is a curve and let Γ be derived from U by normalization with reference to k; then Γ is a complete non-singular curve, which corresponds to U by a birational correspondence T over k. Let $M \times M^*$ be the generic point of T over k with the projection M on U; and let $M^{*\prime}$ be a specialization of M^* over the specialization $X(M) \to Y$, $M \to M'$ with reference to k. If we interchange the second and the third factors of the product $U \times \Gamma \times V \times V$, its subvariety $T \times \Delta$ corresponds biregularly to a birational correspondence Z between $U \times V$ and $\Gamma \times V$. Since Z is biregular along every component of X, the intersection-product $Z \cdot (X \times \Gamma \times V)$ is defined; we put $X^* = \mathrm{pr}_{\Gamma \times V}(Z \cdot (X \times \Gamma \times V))$. Then it can be readily verified that we have

$$X^*(M^*) = \mathrm{pr}_V (X^* \cdot (M^* \times V)) = X(M).$$

Moreover since every component of X^* has the projection Γ on Γ, the intersection-product $X^* \cdot (M^{*\prime} \times V)$ is defined. Therefore by the previous

[6] As in Bourbaki we shall denote by Z the ring of rational integers and by Q its field of quotients.

[7] Cf. S. Lefschetz, *Topology*, American Mathematical Society Colloquium Publication no. 12 (1930).

proposition, its algebraic projection $X^*(M^{*\prime})$ on V is the *uniquely determined* specialization of $X(M)$ over $M^* \to M^{*\prime}$ with reference to k. Since Y is also the specialization of $X(M)$ over the same specialization, we must have $Y = X^*(M^{*\prime})$. Hence Y and $X(M)$ are continuously equivalent.

Now we shall prove that $X(M)$ and $X(M')$ are continuously equivalent, whenever M and M' are generic points of U^s over k. Since this is true for $s = 1$, we can use induction on s assuming $s \geqq 2$. By taking another generic point of U over $k(M, M')$, if necessary, we may assume that M and M' are already independent over k. Let U, (x) and (x') be a set of representatives of U, M and M' in the same ambient space S^m; let $(u^\alpha) = (u_1{}^\alpha, \cdots, u_m{}^\alpha)$ $(\alpha = 0, 1)$ be two sets of $2m$ independent variables over $k(x, x')$ and put

$$v^0 = -\sum_{i=1}^{m} u_i{}^0 x_i, \qquad v^1 = -\sum_{i=1}^{m} u_i{}^1 x'_i.$$

Then the equations $\sum_{i=1}^{m} u_i{}^\alpha X_i + v^\alpha = 0$ define two linear varieties $H_\alpha{}^{m-1}$ in S^m such that the intersection-products $W_\alpha = U \cdot H_\alpha$ are (absolutely irreducible) varieties. Moreover the intersection-product $W = U \cdot H_0 \cdot H_1$ is, at least, a prime rational U-cycle over $k(u^0, v^0, u^1, v^1)$; and a generic point N of W over this field is a generic point of W_α over $k(u^\alpha, v^\alpha)$ for $\alpha = 0, 1$. Now let $W_\alpha{}^{s-1}$ and N be the subvarieties of U^s, which have the representatives W_α and N in U; since the components of $X \cap (W_\alpha \times V)$, which have the projection W_α on U, are all proper, we can form a $(W_\alpha \times V)$-cycle X_α by such components. Then we have

$$X_0(M) = \mathrm{pr}_V\left((X_0 \cdot (M \times V))_{W_0 \times V}\right) = X(M),$$
$$X_0(N) = X(N) = X_1(N), \qquad X_1(M') = X(M');$$

therefore by the induction assumption, $X(M)$ and $X(M')$ are continuously equivalent.

Finally, even if M' is not a generic point of U over k, we may assume that Y is the uniquely determined specialization of $X(M)$ over $M \to M'$ with reference to k. Otherwise let M^* and $M^{*\prime}$ be the Chow points of $X(M)$ and Y; since M^* is rational over $k(M)$, it has a locus U^* over k. Moreover, since $X(M)$ is rational over $k(M^*)$, there exists a rational $(U^* \times V)$-cycle X^* over k, every component of which has the projection U^* on U^*, such that

$$X^*(M^*) = \mathrm{pr}_V\, (X^* \cdot (M^* \times V)) = X(M).$$

Then we have only to replace U, X and M' by U^*, X^* and $M^{*\prime}$. Now by Prop. 7 of (F), Appendix II, there exists a generic specialization N of M

over k and an extension K of k such that N has a locus U_0 of dimension 1 over K containing M'. As before let X_0 be the $(\mathsf{U}_0 \times \mathsf{V})$-cycle composed of those components of $\mathsf{X} \cap (\mathsf{U}_0 \times \mathsf{V})$, which have the projection U_0 on U; we then have

$$\mathsf{X}_0(\mathsf{N}) = \mathrm{pr}_{\mathsf{V}}\ ((\mathsf{X}_0 \cdot (\mathsf{N} \times \mathsf{V}))_{\mathsf{U}_0 \times \mathsf{V}}) = \mathsf{X}(\mathsf{N}).$$

By assumption, Y is also the specialization of $\mathsf{X}(\mathsf{N})$ over $\mathsf{N} \rightarrow \mathsf{M}'$ with reference to K. Therefore since U_0 is a curve, Y and $\mathsf{X}(\mathsf{N})$, and so, by our second step, Y and $\mathsf{X}(\mathsf{M})$ are continuously equivalent.

§ II. Continuous Family of Curves in V.

4. Let W^1 be a generic 1-section of V over a field k and let M be the Chow point of W^1. Since W^1 is defined over a purely transcendental extension of k, $k(\mathsf{M})$ is regular over k and M has a locus U over k. Moreover there exists a rational $(\mathsf{U} \times \mathsf{V})$-cycle C over k, every component of which has the projection U on U, such that

$$\mathsf{C}(\mathsf{M}) = \mathrm{pr}_{\mathsf{V}}\ (\mathsf{C} \cdot (\mathsf{M} \times \mathsf{V})) = \mathsf{W}^1.$$

Since W^1 is a variety, C must also be a variety; we shall study the properties of this continuous family in V with a "generic curve" $\mathsf{C}(\mathsf{M})$ over k.

A differential form or a differential Φ on V, which can be expressed in the form $\Phi = \Sigma F_{i_1 \ldots i_s}(z) dz_{i_1} \wedge \cdots \wedge dz_{i_s}$, where the coefficients $F_{i_1 \ldots i_s}(z)$ and the "uniformizing parameters" $z_1, \cdots, z_d$ are (algebraic) functions on V^d, will be called an *algebraic differential of degree s on* V. Here $\wedge$ denotes Grassmann multiplication. An algebraic differential of degree 1 on V will be called a Picard differential on V. It can be proved easily by the last corollary in (C) that in the above definition we have only to assume that the coefficients $F_{i_1 \ldots i_s}(z)$ are meromorphic in $z_1, \cdots, z_d$ for every choice of $z_1, \cdots, z_d$. If $F_{i_1 \ldots i_s}(z)$ are always regular analytic in $z_1, \cdots, z_d$, Φ is said to be *of the first kind.* It is then a complex harmonic form in the sense of Hodge on the compact manifold V with a "natural" Kählerian metric.[8]

THEOREM 2. *The set of linearly independent Picard differentials of the first kind on* V *remains linearly independent on each generic curve* $\mathsf{C}(\mathsf{M})$ *over k.*

---

[8] Cf. W. V. D. Hodge, *The theory and applications of harmonic integrals*, Cambridge University Press (1941); Chap. IV. We shall cite this book as (H). The theorem which is stated above is implicite in (H), but can be proved as follows. By the special nature of the Kählerian metric, the "dual form" of every algebraic differential of the first kind is exact. Therefore if Φ is such a form, $d\Phi$ and its dual form are both exact. This shows precisely that $d\Phi$ is harmonic. Since it is at the same time a derived form, we have $d\Phi = 0$.

We shall give the "analytic" proof of Theorem 2 at the end of this section, and at present, we shall derive some of its consequences. Let p be the genus of $\mathsf{C}(\mathsf{M})$; since the generic curves are the transforms of one another in the sense of (F), Chap. VII, § 2, p does not depend on M. Moreover by Theorem 2, we can find a base of the Picard differentials of the first kind on V

$$(\Phi) = (\Phi_1, \Phi_2, \cdots, \Phi_q) \qquad (p \geqq q \geqq 0);$$

the number q is called the *irregularity* of V.

We shall denote by $\mathfrak{H}^r(\mathsf{V}, Z)$ or $\mathfrak{H}^r(\mathsf{V}, Q)$ the r-dimensional homology group of V modulo Z or Q, and by $R^r(\mathsf{V})$ the rank of $\mathfrak{H}^r(\mathsf{V}, Q)$. Hodge[9] gave an elegant proof of the classical theorem that $R^1(\mathsf{V}) = 2q$. Now let $\gamma_1, \cdots, \gamma_{2q}, \delta_1, \cdots, \delta_t$ be a base of $\mathfrak{H}^1(\mathsf{V}, Z)$, where $(\gamma) = (\gamma_1, \gamma_2, \cdots, \gamma_{2q})$ is a base of the 1-dimensional Betti group $\mathfrak{B}(\mathsf{V})$ of V modulo Z and where $\delta_1, \delta_2, \cdots, \delta_t$ are torsion cycles. Although the torsion cycles are absent if V is a curve, they actually appear in the general case. It is clear that the period matrix $\omega = (\int \gamma_i \Phi_\alpha)$ of the "Picard integrals" $\int \Phi_\alpha$ $(1 \leqq \alpha \leqq q)$ along the 1-cycles γ_i $(1 \leqq i \leqq 2q)$, is determined by V up to an "isomorphism" $\omega \to \Lambda \omega L$, where Λ is a non-singular K-matrix of degree q, and L a unimodular Z-matrix of degree $2q$.

On the other hand, let $(\alpha) = (\alpha_1, \alpha_2, \cdots, \alpha_{2p})$ be a base of $\mathfrak{H}^1(\mathsf{C}(\mathsf{M}), Z)$. Then there exists a uniquely determined Z-matrix A of type $(2q, 2p)$ such that $(\alpha) \sim (\gamma)A$ (mod. Q). In this paper the skew-symmetric Z-matrix $E_{\mathsf{M}} = A^t I_\alpha^{-1t} A$ plays an important rôle,[10] where I_α means the unimodular Z-matrix of Kronecker indices $I(\alpha_i, \alpha_j; \mathsf{C}(\mathsf{M}))$ on $\mathsf{C}(\mathsf{M})$. It can be readily verified that E_{M} does not depend on the choice of the base (α), if the base (γ) and the generic curve $\mathsf{C}(\mathsf{M})$ over k are given. We shall see later that E_{M} does not depend on M.

5. Applying the celebrated theorem of Poincaré in the theory of "reducible integrals"[11] to the above circumstances, we can find a base (β) of $\mathfrak{H}^1(\mathsf{C}(\mathsf{M}), Q)$, which is composed of "invariant cycles" $\beta_i \sim \gamma_i$ (mod. Q)

---

[9] Cf. (H), § 49. See also his original proof: "Harmonic integrals associated with algebraic varieties," *Proceedings of the London Mathematical Society*, vol. 39 (1935).

[10] Let M be a matrix with any coefficients, then ${}^tM$ means the transposed matrix of M. Moreover if M is a K-matrix, we shall denote by $\overline{M}$ the image of M by the involutive automorphism of K over the real field.

[11] Cf. O. Zariski, *Algebraic surfaces*, Ergebnisse der Mathematik (1935), Chap. VII, § 6.

$(i \leqq 2q)$, and of "vanishing cycles" $\beta_i \sim 0$ (in V mod. Q) $(i > 2q)$, such that if we define a square Q-matrix B of degree $2p$ by

$$(\alpha) \sim (\beta)B \qquad (\text{on } \mathsf{C}(\mathsf{M}) \text{ mod. } Q),$$

$$B^t I_\alpha^{-1t} B = \begin{pmatrix} E_{\mathsf{M}} & 0 \\ 0 & \end{pmatrix}.$$

Moreover we can find a "complementary set" of Abelian integrals $\int \Phi_{q+1}, \cdots, \int \Phi_p$ on $\mathsf{C}(\mathsf{M})$ such that the period matrix Ω_β of $\int \Phi_\alpha$ $(1 \leqq \alpha \leqq p)$ along the 1-cycles β_i $(1 \leqq i \leqq 2p)$ splits in the following form:

$$\Omega_\beta = \begin{pmatrix} \omega & 0 \\ 0 & \end{pmatrix}.$$

It follows that ${}^tE_{\mathsf{M}} = -E_{\mathsf{M}}$ is a "principal matrix" of the "Riemann matrix" ω.

Now the $2q$ columns of ω generate a discrete subgroup $[\omega]$ of rank $2q$ in the complex vector space S^q, and therefore the factor group $S^q/[\omega]$ is a complex toroid of (complex) dimension q. Moreover, since ω is shown to be a Riemann matrix, we see that this complex toroid is mapped in a one-to-one way to a non-singular variety A^q in a suitable projective space by means of theta functions $\Theta(v) = \Theta(v_1, \cdots, v_q)$, belonging to $[\omega]$.[12] It then follows from (C) that A^q is an abelian variety in the sense of Weil[13]; following him we shall call A the *Albanese variety attached to* V. Since the period matrix ω is determined up to an isomorphism, A is determined up to an isomorphism in the sense of (V) by our projective model V. We shall see later that A is attached to V in a birationally invariant way.

6. Now we shall prove our Theorem 2. There exists, by (H), Chap. IV, a harmonic 2-form ν in L^n such that $\int_{\mathsf{D}} \nu = 1$, where D is any projective 1-space in L^n. ν induces a similar form on V, which we shall denote by the same letter; and we shall show that ν and the generic $(d-1)$-section $\mathsf{W} = \mathsf{V} \cdot \mathsf{H}$ of V^d are homologous in the sense of de Rham.

LEMMA. *Let Z be any topological 2-cycle in* V. *Then* $\int_Z \nu = I(Z, \mathsf{W}; \mathsf{V})$.

---

[12] Cf. S. Lefschetz, "On certain numerical invariants of algebraic varieties with applications to abelian verieties," *Transactions of the American Mathematical Society*, vol. 22 (1921).

[13] A. Weil, "Variétés abeliennes et courbes algébriques," *Actualité Scientifiques et Industrielles*, no. 1064 (1948). We cite this book as (V).

Proof. Since we have $\int_Z \nu = I(Z, \mathsf{H})$, it is sufficient to show that the equality $Z \cdot \mathsf{H} = (Z \cdot (\mathsf{V} \cdot \mathsf{H}))_{\mathsf{V}}$. Thereby we may assume that Z is at least "differentiable" in the neighborhoods of the points in $Z \cap \mathsf{H}$ and that Z is transversal to H at these points. Then our formula surely holds locally, hence in the large.[14]

Now we can prove the following remarkable fact, which was discovered by Lefschetz.[15]

PROPOSITION 3. *Let $Z_1, \cdots, Z_R$ $(R = R^{2d-r}(\mathsf{V}))$ be a base of $\mathfrak{H}^{2d-r}(\mathsf{V}, Q)$ $(0 \leqq r \leqq d)$. Then the r-cycles*

$$\Gamma_i = (Z_i \cdot \mathsf{W})_{\mathsf{V}} \qquad (1 \leqq i \leqq R)$$

on the generic r-section W of V^d are independent as cycles in V modulo Q.

Proof. By Hodge's existence theorem[16] there exists a set of harmonic r-forms $\phi_1, \cdots, \phi_R$ in V such that $\int_\Gamma \phi_i = I(Z_i, \Gamma; \mathsf{V})$ $(1 \leqq i \leqq R)$ for every topological r-cycle Γ. Consider the harmonic $2(d-r)$-form $\nu_{d-r} = \nu \wedge \cdots \wedge \nu$ $((d-r)$-factors). Then by Theorem III and by Lemma C, § 42 in (H), we conclude that the harmonic $(2r-r)$-forms $\phi_i \wedge \nu_{d-r}$ $(1 \leqq i \leqq R)$ are linearly independent. By de Rham's theorem, however, $\int_Z \phi_i \wedge \nu_{d-r} = I(\Gamma_i, Z; \mathsf{V})$ $(1 \leqq i \leqq R)$ for every topological $(2d-r)$-cycle Z; our proposition follows from this fact.

COROLLARY. *Every topological r-cycle in V^d $(0 \leqq r \leqq d)$ is homologous to some cycle on the generic s-section W^s of V $(s \geqq r)$* modulo Q.

Theorem 2 is now a simple consequence of this corollary in the case of $r = s = 1$. However a purely algebraic proof for this theorem is greatly to be desired.

§ III. Poincaré Family and Picard Variety Attached to V.

7. We shall start from the following q "Abelian sums"

$$\sum_{i=1}^{q} \int^{\mathsf{P}_i} \Phi_\alpha = v_\alpha \qquad (1 \leqq \alpha \leqq q),$$

---

[14] See loc. cit. 7.

[15] S. Lefshetz, *L'analysis situs et la géométrie algébrique*, Paris (1924), Chap. V.

[16] Cf. K. Kodaira, "Harmonic fields in Riemannian manifolds," *Annals of Mathematics*, vol. 50 (1949).

where $\mathsf{P}_1, \cdots, \mathsf{P}_q$ are q points in V. If we restrict these points to the generic curve $\mathsf{C}(\mathsf{M})$ over k, we get an analytic mapping from the product $\tilde{\mathsf{C}}(\mathsf{M}) = \mathsf{C}(\mathsf{M}) \times \cdots \times \mathsf{C}(\mathsf{M})$ (q-factors) into the complex toroid $S^q/[\omega]$, which we shall denote by $\boldsymbol{\omega}$, in a natural way.

PROPOSITION 4. *The image* $\tilde{\boldsymbol{\omega}}$ *of* $\tilde{\mathsf{C}}(\mathsf{M})$ *covers the whole of* $\boldsymbol{\omega}$, *and covers in general exactly* $q!\,\epsilon_{\mathsf{M}}$*-times; where* ϵ_{M} *denotes the "Pfaffian" of* E_{M}.

Proof. Let $\mathsf{P}_1, \cdots, \mathsf{P}_q$ be a set of q distinct points on $\mathsf{C}(\mathsf{M})$; we shall show that the Jacobian of our mapping does not vanish there identically. In fact let $x_1, \cdots, x_q$ be the uniformizing parameters of $\mathsf{C}(\mathsf{M})$ at $\mathsf{P}_1, \cdots, \mathsf{P}_q$ respectively; then in the contrary case we have $\det|\Phi_\alpha(x_\beta)| = 0$, identically in $x_1, \cdots, x_q$. This, however, would contradict the linear independence of $\Phi_1, \cdots, \Phi_q$ on $\mathsf{C}(\mathsf{M})$. Therefore if we regard $\tilde{\boldsymbol{\omega}}$ as a compact connected "covering variety" of $\boldsymbol{\omega}$, its projection on $\boldsymbol{\omega}$ contains an open set of $\boldsymbol{\omega}$. Since the projection $\mathfrak{J}$ of the points of $\tilde{\boldsymbol{\omega}}$ on $\boldsymbol{\omega}$, at which the Jacobian vanishes, forms an analytic bunch of $(q-1)$ (complex) dimensions (at most) in $\boldsymbol{\omega}$, its complement $\boldsymbol{\omega} - \mathfrak{J}$ is connected. Our proposition follows from this fact except for the exact value of the covering; and this crucial result can be proved by Wirtinger's method [17] as follows.

Let $m(\boldsymbol{\omega})$ and $m(\tilde{\boldsymbol{\omega}})$ be the "volumes" of $\boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}$ respectively; and put

$$\int^{\mathsf{P}} \Phi_\alpha = y_{2\alpha-1} + i y_{2\alpha} \qquad (1 \leqq \alpha \leqq q).$$

Then it can be readily seen that

$$m(\tilde{\boldsymbol{\omega}}) = \sum_{i<j} \operatorname{sgn}\begin{pmatrix} i_1 j_1 \cdots i_q j_q \\ 1 \cdots\cdots 2q \end{pmatrix} \prod_{h=1}^{q} \oint y_{i_h} dy_{j_h},$$

where $\oint$ means the "boundary integral" on the Riemann surface of $\mathsf{C}(\mathsf{M})$.

Therefore we have

$$m(\tilde{\boldsymbol{\omega}}) = m(\boldsymbol{\omega}) \sum_{i<j} \operatorname{sgn}\begin{pmatrix} i_1 j_1 \cdots i_q j_q \\ 1 \cdots\cdots 2q \end{pmatrix} e_{i_1 j_1} \cdots e_{i_q j_q},$$

where e_{ij} $(1 \leqq i, j \leqq 2q)$ are the coefficients of E_{M}; hence $m(\tilde{\boldsymbol{\omega}}) : m(\boldsymbol{\omega}) = q!(\det|E_{\mathsf{M}}|)^{\frac{1}{2}}$, which completes the proof.

---

[17] W. Wirtinger, "Zur Theorie der $2n$-fach periodischen Funktionen," *Monatshefte für Mathematik*, Bd. 7 (1896).

Now if we introduce a function Φ on V by writing

$$\Phi(P) = \Theta(\int^{P} \Phi_1, \cdots, \int^{P} \Phi_q),$$

the function $\tilde{\Phi}$ on $\tilde{V} = V \times \cdots \times V$ (q-factors)

$$\tilde{\Phi}(P_1, \cdots, P_q) = \Theta(\sum_{i=1}^{q} \int^{P_i} \Phi_1, \cdots, \sum_{i=1}^{q} \int^{P_i} \Phi_q),$$

can be written in the form

$$\tilde{\Phi}(P_1, \cdots, P_q) = \sum_{i=1}^{q} \Phi(P_i);$$

where the right side refers to the group variety A. The fact that Φ and hence also $\tilde{\Phi}$ are (algebraic) functions follows from (C) Theorem 7. If M is a generic point of U over a common field of definition K of C, Φ, A and of the composition function in A, the functions Φ and $\tilde{\Phi}$ induce the function ϕ and $\tilde{\phi}$ on $C(M)$ and $C(M)$ respectively, such that their graphs are related as follows:

$$\Gamma_{\Phi} \cdot (C(M) \times A) = \Gamma_{\phi}, \qquad \Gamma_{\tilde{\Phi}} \cdot (\tilde{C}(M) \times A) = \Gamma_{\tilde{\phi}}.$$

Hence, using Proposition 4, we can obtain the following result, which is well known if V is a curve.

THEOREM 3. *There exists a function Φ on V with values in the Albanese variety A^q attached to V. Moreover on each generic curve $C(M)$ over K, Φ induces a function ϕ such that if $P_1, \cdots, P_q$ are q independent generic points of $C(M)$ over $K(M)$, then the point $z = \sum_{i=1}^{q} \phi(P_i)$, is a generic point of A over $K(M)$ and satisfies*

$$[K(M)(P_1, \cdots, P_q)_s : K(M)(z)] = \epsilon_M;$$

where $K(M)(P_1, \cdots, P_q)_s$ is the invariant subfield of $K(M)(P_1, \cdots, P_q)$ by $q!$ permutations of $P_1, \cdots, P_q$.

We conclude from this theorem that ϵ_M does not depend on the choice of $C(M)$; hence we may write it as ϵ.

Now the point $z \times P_1$, say, in the product $A \times V$ is rational over $K(M, P_1, \cdots, P_q)$; hence $K(z, P_1)$ is regular over K and $z \times P_1$ has a locus X over K. Moreover the V-cycle $X(z) = \mathrm{pr}_V (X \cdot (z \times V))$ is prime rational over $K(z)$ and has a non-empty intersection with the generic curve $C(M)$ over $K(z)$. Since P_1 is algebraic over $K(M, z)$, $X(z)$ is a prime rational V-divisor over $K(z)$; and the intersection-product $X(z) \cdot C(M)$ is a prime rational $\mathsf{C}(\mathsf{M})$-divisor over $K(\mathsf{M}, z)$. On the other hand, the $\mathsf{C}(\mathsf{M})$-divisor $\mathrm{pr}_{\mathsf{C}(\mathsf{M})}(\bar{\Phi}^{-1}(z))$ consists of $q\epsilon$ conjugate points of P_1 over $K(\mathsf{M}, z)$, each being repeated $(q-1)$!-times; therefore we have

$$\mathsf{X}(z) \cdot \mathsf{C}(\mathsf{M}) = 1/(q-1)!\, \mathrm{pr}_{\mathsf{C}(\mathsf{M})}(\bar{\Phi}^{-1}(z)).$$

For historical reasons we shall attach the name of *Poincaré family* to this continuous family X. If Y is a member of the Poincaré family and if z' is the specialization of z over the specialization $\mathsf{X}(z) \to \mathsf{Y}$ with reference to K, we shall write $\mathsf{Y} = \mathsf{X}(z')$. Since such $\mathsf{X}(z')$ are continuously equivalent in the sense of § I, 2, they are homologous to each other modulo Z.

8. If a V-divisor Y is homologous to 0 modulo Q, there exists a Picard differential of the "third kind" Ψ on V with the "residue" Y. This is known as Lefschetz's theorem [18] and proved elegantly by Weil.[19] We may assume that the real part of the integral of Ψ is one-valued on V, for otherwise we have only to add a suitable linear combination of Φ_α $(1 \leqq \alpha \leqq q)$ to Ψ. Now let δ be any torsion 1-cycle in V such that $m\delta \sim 0$ (mod. Z); let C be a 2-chain over Z with the boundary $m\delta$. Then it can be readily seen from Stokes theorem that

$$\int_\delta \Psi = 2\pi i/mI(C, \mathsf{Y}; \mathsf{V}) = 2\pi i L(\delta, \mathsf{Y}; \mathsf{V}),$$

where $L(\delta, \mathsf{Y}; \mathsf{V})$ denotes the linking coefficient of δ and Y on V. Therefore if we put $F(\mathsf{P}) = \exp(\int^{\mathsf{P}} \Psi)$, F is meromorphic on V; and if we continue F analytically along a continuous closed path Γ in V, F is multiplied by a constant factor $\chi_{\mathsf{Y}}(\Gamma)$ of absolute value 1, which depends on the homology class of Γ modulo Z only. Now by the preparation theorem of Weierstrass,[20] we can define the V-divisor (F) for such a "multiplicative function" F on V, and we have $(F) = \mathsf{Y}$. Since the Picard integral of the first kind with pure imaginary periods must be a constant (as follows from the fact that ω is a Riemann matrix), F is uniquely determined up to a constant factor by Y. Therefore each Y determines *uniquely* a character $\chi_{\mathsf{Y}}(\Gamma)$ of the discrete group $\mathfrak{H}^1(\mathsf{V}, Z)$. For a torsion cycle δ, we have

$$\chi_{\mathsf{Y}}(\delta) = \exp(2\pi i L(\delta, \mathsf{Y}; \mathsf{V}));$$

---

[18] Cf. loc. cit. 15, note I.

[19] A. Weil, "Sur la théorie des formes différentielles attaché à une variété analytique complexe," *Comment. Math. Helv.*, vol. 20 (1947).

[20] Cf. S. Bochner and W. T. Martin, *Several complex variables*, Princeton University Press (1948), Chap. IX.

hence $\chi_{\mathsf{Y}}(\delta)=1$, whenever Y is homologous to 0 modulo Z. Such a Y thus induces a character of the Betti group $\mathfrak{B}(\mathsf{V})$; and the difference $\mathsf{Y}=\mathsf{X}(z)-\mathsf{X}(z')$ is surely such a V-divisor. In this case we shall obtain an explicit formula of the character

$$(\chi_{\mathsf{Y}}(\gamma))=(\chi_{\mathsf{Y}}(\gamma_1),\cdots,\chi_{\mathsf{Y}}(\gamma_{2q})).$$

9. We may first assume, by taking a suitable extension of K if necesary, that $\mathsf{X}(z')$ is rational over K. Let L be an extension of K over which $\mathsf{X}(z)$ is rational and let M and w be independent generic points of U and A over L. Then the intersection-products $\mathsf{X}(z)\cdot\mathsf{C}(\mathsf{M})$ and $\mathsf{X}(w)\cdot\mathsf{C}(\mathsf{M})$ are defined, and by definition the former is a specialization of the latter over $w\to z$ with reference to $L(\mathsf{M})$. On the other hand we have

$$\mathsf{X}(w)\cdot\mathsf{C}(\mathsf{M})=\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\mathsf{Q}_{ij}$$

with $\sum_{j=1}^{q}\phi(\mathsf{Q}_{ij})=w$ $(1\leqq i\leqq\epsilon)$ by Theorem 3. Hence we may write $\mathsf{X}(z)\cdot\mathsf{C}(\mathsf{M})$ in the form

$$\mathsf{X}(z)\cdot\mathsf{C}(\mathsf{M})=\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\mathsf{P}_{ij},$$

where $(\cdots,\mathsf{P}_{ij},\cdots)$ is the specialization of $(\cdots,\mathsf{Q}_{ij},\cdots)$ over $w\to z$ with reference to $L(\mathsf{M})$. Since the (algebraic) function $w=\tilde{\Phi}(\mathsf{Q}_{i1},\cdots,\mathsf{Q}_{iq})$ is "defined" at $(\mathsf{P}_{i1},\cdots,\mathsf{P}_{iq})$, the "value" of $\tilde{\Phi}$ at this point is the *uniquely determined* specialization of w over $(\mathsf{Q}_{i1},\cdots,\mathsf{Q}_{iq})\to(\mathsf{P}_{i1},\cdots,\mathsf{P}_{iq})$ with reference to $L(\mathsf{M})$. We therefore have

$$z=\tilde{\Phi}(\mathsf{P}_{i1},\cdots,\mathsf{P}_{iq})=\Theta\Big(\sum_{j=1}^{q}\int^{\mathsf{P}_{ij}}\Phi_1,\cdots,\sum_{j=1}^{q}\int^{\mathsf{P}_{ij}}\Phi_q\Big)\qquad(1\leqq i\leqq\epsilon).$$

Similarly $\mathsf{X}(z')\cdot\mathsf{C}(\mathsf{M})$ can be written in the form

$$\mathsf{X}(z')\cdot\mathsf{C}(\mathsf{M})=\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\mathsf{P}'_{ij},$$

and we have

$$z'=\Theta\Big(\sum_{j=1}^{q}\int^{\mathsf{P}'_{ij}}\Phi_1,\cdots,\sum_{j=1}^{q}\int^{\mathsf{P}'_{ij}}\Phi_q\Big)\qquad(1\leqq i\leqq\epsilon).$$

Therefore if $z-z'$ has the "coordinates" (v), then

$$\sum_{i=1}^{\epsilon}\sum_{j=1}^{q}\int_{\mathsf{P}'_{ij}}^{\mathsf{P}_{ij}}(\Phi)=\epsilon\cdot(v),$$

where $(\Phi)=(\Phi_1,\cdots,\Phi_q)$.

Now the multiplicative function F on V induces a similar function f on $\mathsf{C}(\mathsf{M})$ such that

$$(f) = (F)\cdot \mathsf{C}(\mathsf{M}) = \mathsf{X}(z)\cdot \mathsf{C}(\mathsf{M}) - \mathsf{X}(z')\cdot \mathsf{C}(\mathsf{M}).$$

But since $(\gamma_1,\cdots,\gamma_{2q},0,\cdots,0)\sim(\beta)\sim(\alpha)\cdot B^{-1}$ (mod. Q), it follows that if e is a natural number which makes eB^{-1} into a Z-matrix, and if $(\phi(\alpha)) = (\phi(\alpha_1),\phi(\alpha_2),\cdots,\phi(\alpha_{2p}))$ is a base of the harmonic 1-forms on $\mathsf{C}(\mathsf{M})$, which is homologous to (α) in the sense of de Rham, then we have

$$(\chi_{\mathsf{Y}}(\gamma)^e\chi_{\mathsf{Y}}(0)^e) = \exp(2\pi i \sum_{ij}\int_{\mathsf{P}'_{ij}}^{\mathsf{P}_{ij}}(\phi(\alpha))(eB^{-1})).^{21}$$

On the other hand there exists a K-matrix X of type $(p,2p)$ such that $(\phi(\alpha)) = (\Phi\bar{\Phi})\begin{pmatrix} X \\ \bar{X}\end{pmatrix}$, where (Φ) has p columns. By integrating this equation, we get $I_\alpha = {}^tB({}^t\Omega_\beta\,{}^t\Omega_\beta)\begin{pmatrix} X \\ \bar{X}\end{pmatrix}$; if we put $\sum_{ij}\int_{\mathsf{P}'_{ij}}^{\mathsf{P}_{ij}}(\Phi) = (v^*)$, we have

$$(\chi_{\mathsf{Y}}(\gamma)^e 1) = \exp(2\pi i (v^*\bar{v}^*)\begin{pmatrix} X \\ \bar{X}\end{pmatrix}(eB^{-1})).$$

Now there exists a real vector (m') with $2p$ columns such that $(v^*) = (m')^t\Omega_\beta$; and we have

$$\begin{aligned}(\chi_{\mathsf{Y}}(\gamma)^e 1) &= \exp(2\pi i e(m')\,(BI_\alpha{}^{-1}{}^tB)^{-1} \\ &= \exp(2\pi i e(m')\begin{pmatrix} \boxed{{}^tE_{\mathsf{M}}{}^{-1}} & 0 \\ 0 & \boxed{} \end{pmatrix}).\end{aligned}$$

Therefore if we denote by (m) a real vector with $2q$ columns such that $(v) = (m)^t\omega$, we have

$$(\chi_{\mathsf{Y}}(\gamma)^e) = \exp(2\pi i e\epsilon(m)^tE_{\mathsf{M}}{}^{-1}).$$

Now we may take as z every generic point of A over $K(\mathsf{M})$; such points are everywher dense in A. However since $(\chi_{\mathsf{Y}}(\gamma))$ and $\exp(2\pi i\epsilon(m)^tE_{\mathsf{M}}{}^{-1})$ both depend continuously on z,[22] and they approach (1), if z approaches z', we have

$$(\chi_{\mathsf{Y}}(\gamma)) = \exp(2\pi i\epsilon(m)^tE^{-1}).$$

---

[21] Cf. H. Weyl, *Die Idee der Riemannschen Fläche*, Berlin (1913), Kap. II, §§ 16-17. See also J. Igusa, "Zur klassischen Theorie der algebraischen Funktionen," *Journal of the Mathematical Society of Japan*, vol. 1 (1948).

[22] That the first character depend continuously upon z can be proved rigorously by Kodaira's generalization of Weyl's formula: Green's formula and meromorphic functions on compact analytic varieties, to appear in the *Canadian Journal of Mathematics*.

Thereby we write E instead of E_{M}, since it does not depend on the choice of $\mathsf{C}(\mathsf{M})$. In fact this follows from the above formula, if we remember that $(\chi_{\mathsf{Y}}(\gamma))$, ϵ and (m) are independent of $\mathsf{C}(\mathsf{M})$.

10. On the other hand if we replace the period matrix ω by $\hat{\omega} = \omega\epsilon^{-1}E$, $[\omega]$ is a subgroup of index ϵ^{2q-2} in the similar group $[\hat{\omega}]$. Moreover since $\hat{\omega}$ is also a Riemann matrix, there exists an abelian variety $\mathfrak{P}$ in a suitable projective space, which is isomorphic with $S^q/[\hat{\omega}]$ as complex toroids, and a homomorphism λ from A to $\mathfrak{P}$ such that $\lambda\Theta(v) = 0$ if and only if (v) is contained in $[\hat{\omega}]$. If we define a real vector (r) with $2q$ columns by $(v) = (r)^t\hat{\omega}$, we have $(\chi_{\mathsf{Y}}(\gamma)) = \exp(2\pi i(r))$. Therefore if we denote by $\mathfrak{G}_l(\mathsf{V})$ the group of V-divisors, which are linearly equivalent to 0 in V, then $\mathsf{X}(z_1) \equiv \mathsf{X}(z_2) \pmod{\mathfrak{G}_l(\mathsf{V})}$ if and only if $\lambda(z_1) = \lambda(z_2)$. This abelian variety $\mathfrak{P}$ is the same as the *Picard variety attached to* V in the sense of the Italian geometers.

It follows from the above results that if a V-divisor Y is homologous to 0 modulo Z, we can find a V-divisor $\mathsf{X}(z) - \mathsf{X}(z')$, which induces the same character on $\mathfrak{B}(\mathsf{V})$ as Y. Then, by the definition of the character, Y is linearly equivalent to $\mathsf{X}(z) - \mathsf{X}(z')$. Since linear equivalence implies continuous equivalence in the sense of § I, 2, and since $\mathsf{X}(z)$ is continuously equivalent to $\mathsf{X}(z')$, by Theorem 1, we see readily that Y is continuously equivalent to 0. As we have remarked in § I, 2, however, continuous equivalence implies the homological equivalence modulo Z; thus we can state

THEOREM 4. *In the case of divisors on the projective model, the continuous equivalence and the homological equivalence* modulo Z *are the same.*[23]

This theorem was first proved by Lefschetz [24] essentially along the same line as above. We shall denote by $\mathfrak{G}_c(\mathsf{V})$ the group of V-divisors, which is

---

[23] If $\mathfrak{G}^r{}_w(\mathsf{V}_a)$ and $\mathfrak{G}^r{}_{w'}(\mathsf{V}_a)$ correspond to the complete variety V_a without multiple point $(r \leqq \dim(\mathsf{V}_a))$ by the "strictest equivalence theories" satisfying (A), (B), (C′), (D), (E), (S) and (A), (D), (E), (S) of (F), Chap. IX, § 7 respectively, we have

$$\mathfrak{G}^{d-1}{}_w(\mathsf{V}) = \mathfrak{G}^{d-1}{}_{w'}(\mathsf{V}) = \mathfrak{G}_c(\mathsf{V}).$$

In fact if $\mathfrak{G}^r{}_h(\mathsf{V}_a)$ corresponds to V_a by the homological equivalence theory over Z, it is an equivalence theory satisfying (A), (B), (C′), (D), (E) and (S); hence

$$\mathfrak{G}^r{}_{w'}(\mathsf{V}_a) \subset \mathfrak{G}^r{}_w(\mathsf{V}_a) \subset \mathfrak{G}^r{}_h(\mathsf{V}_a)$$

for every V_a. Moreover by the same arguments as in § I, 2, we have

$$\mathfrak{G}_c(\mathsf{V}) \subset \mathfrak{G}^{d-1}{}_{w'}(\mathsf{V});$$

and $\mathfrak{G}^{d-1}(\mathsf{V}) = \mathfrak{G}_c(\mathsf{V})$ by Theorem 4.

[24] Cf. loc. cit. 15, Chap. IV.

defined by the continuous equivalence on V, and we shall resume our main results, which include the classical "inversion theorem" of Jacobi, as follows.

THEOREM 5. *We can attach two abelian varieties, the Albanese variety* A *and the Picard variety* $\mathfrak{P}$, *to every non-singular projective model* V. $\mathfrak{P}$ *is obtained by a homomorphism from* A, *which corresponds to the division of the period matrix of the Picard integrals of the first kind on* V *by one of its principal matrices. Moreover,* $\mathfrak{P}$ *is isomorphic with the factor group* $\mathfrak{G}_c(\mathsf{V})/\mathfrak{G}_l(\mathsf{V})$, *and this is dually paired with the Betti group* $\mathfrak{B}(\mathsf{V})$ *by the multiplication rule*

$$\left.\begin{array}{l}\mathfrak{G}_c(\mathsf{V})\ni \mathsf{Y}\\ \mathfrak{B}(\mathsf{V})\ni\gamma\end{array}\right\}\ \mathsf{Y}\cdot\gamma=\chi_{\mathsf{Y}}(\gamma).$$

In addition every element of $\mathfrak{G}_c(\mathsf{V})/\mathfrak{G}_l(\mathsf{V})$ has a representative of the form $\mathsf{X}(z)-\mathsf{X}(z')$ with $\deg.\mathsf{X}(z)=\deg.\mathsf{X}(z')=q\epsilon$.

On the other hand if we introduce a square *real* matrix J of degree $2q$ by $i\omega=\omega J$, then an isomorphism $\omega\rightarrow\Lambda\omega L$ induces an isomorphism $J\rightarrow L^{-1}JL$. Conversely ω is determined up to a special isomorphism $\omega\rightarrow\Lambda\omega$ by J. This matrix J is explicitly introduced in the theory of complex toroids by Weil[25]; hence it may be called the *Weil matrix attached to* ω. A real matrix J is a Weil matrix if and only if $J^2=-1$ and if there exists a skew-symmetric Q-matrix E such that EJ is symmetric and definite. On the other hand a Q-matrix S satisfies the relation $\omega S^t\omega=0$ if and only if

$$\begin{pmatrix} i & 0\\ 0 & -i\end{pmatrix}\text{ commutes with }\begin{pmatrix}\omega\\ \bar\omega\end{pmatrix}S({}^t\bar\omega\,{}^t\omega);$$

hence if and only if S satisfies the equation $JS^tJ=S$. In particular we have $JE^tJ=E$. It follows therefore that

$$i\hat\omega=i\omega\epsilon^{-1}E=\omega J\epsilon^{-1}E=\omega\epsilon^{-1}E^tJ^{-1}=\hat\omega^tJ^{-1}.$$

Thus we get the following result.

COROLLARY. *If* J *is the Weil matrix attached to* ω, *then* ${}^tJ^{-1}$ *is the Weil matrix attached to* $\hat\omega$.

Now we shall consider the Picard variety of an abelian variety. First of all, if V^d is an abelian variety, the Albanese variety A^q attached to V is isomorphic with V. In fact there exists a K-matrix ω of type $(d,2d)$ such that the kernel of the homomorphism h from the universal covering group S^d

[25] A. Weil, "Théorèmes fondamentaux de la théorie des fonctions thêta," *Seminaire Bourbaki* (Mai, 1949).

of V^d onto V coincides with $[\omega]$. Then if we denote by $(v_1, \cdots, v_d)$ the complex coordinates in S^d, their differentials $dv_\alpha = dv_\alpha(h^{-1}(\mathsf{P}))$ $(1 \leqq \alpha \leqq d)$ are linearly independent Picard differentials of the first kind on V, which form a base of such differentials since $R^1(\mathsf{V}) = 2d$. Moreover, since they have the period matrix ω (along a set of suitable 1-cycles in V), A is isomorphic with $S^d/[\omega]$, hence with V. It follows from this fact and from the above corollary that the Picard variety $\mathfrak{P}$ attached to the given variety V is also attached to the Albanese variety A of V as its Picard variety. Since the operation $J \to {}^tJ^{-1}$ in the same corollary is involutive, we get the following duality of Weil: *The Picard variety of an abelian variety V is not V, but another abelian variety V'; the Picard variety of V' is then V, and there is a kind of duality between them.*

The following figure shows the relations of various varieties, which appear in our theory.

$$\begin{tikzcd}
 & \mathsf{A} \text{ (its own Albanese variety)} \arrow[dd, shift left, "\text{Picard variety}"] \\
\mathsf{V} \arrow[ur, "\text{Albanese variety}"] \arrow[dr, "\text{Picard variety}"'] & \\
 & \mathfrak{P} \text{ (its own Albanese variety)} \arrow[uu, shift left]
\end{tikzcd}$$

11. Until now we have fixed our projective model V; we shall discuss therefore in what manner our theory depends on the choice of it. Let V' be another non-singular projective model, which is equivalent to V by a birational correspondence T. Then there exists a bunch $\mathfrak{B}$ of $(d-1)$ (complex) dimension (at most) on V^d such that T is biregular at every point of $\mathsf{V} - \mathfrak{B}$. Moreover there exist a bunch $\mathfrak{F}$ of (complex) dimension at most $(d-2)$ on V such that the projection from T to V is regular at every point of $\mathsf{V} - \mathfrak{F}$. Therefore T induces an isomorphism of the Poincaré group of V onto that of V'. This fact was remarked explicitly for the first time by Ehresmann on the last page of his thesis.

On the other hand let $\Phi'(\mathsf{P}')$ be any algebraic differental of the first kind on V'. Then the differential $\Phi(\mathsf{P}) = \Phi'(\mathsf{T}(\mathsf{P}))$ on V is also of the first kind. This was proved (explicitly) by Kähler [26]; and the main idea is as follows. If Φ is not of the first kind, it has at least one "pole" Y on V; but since $\mathsf{T}(\mathsf{P})$ is regular along Y, $\Phi'(\mathsf{T}(\mathsf{P}))$ is finite along Y, a contradiction.

---

[26] E. Kähler, "Forme differentiali e funzioni algebriche," *Mem. Accad. Ital. Mat.*, vol. 3 (1932). Cf. also S. Koizumi, "On the differential forms of the first kind on algebraic varieties," *Journal of the Mathematical Society of Japan*, vol. I (1949).

Therefore if ω is the period matrix of the Picard integrals of the first kind on V, it is also a similar matrix for V'. We have thus arrived at the following result.

THEOREM 6. *If A and A' are the Albanese varieties attached to V and V' respectively, then they are isomorphic.*

A simple consequence of this theorem and corollary to Theorem 5 is

THEOREM 7. *If $\mathfrak{P}$ and $\mathfrak{P}'$ are the Picard varieties attached to V and V' respectively, then they are isomorphic.*

We add two Appendices in which, it is hoped, a fairly complete theory of divisors on a non-singular projective model will be established.

Apendix I. Numerical Equivalence and the Lefschetz Number.

We shall first obtain a duality theorem including the torsion cycles. By a duality of Poincaré, the 1-dimensional torsion group of V^d is isomorphic with the $(2d-2)$-dimensional torsion group of V. This duality is based on the fact that we can select the "dual bases" $\delta_1, \cdots, \delta_t$ and $\mathsf{Y}_1, \cdots, \mathsf{Y}_t$ of the torsion groups of 1 and $(2d-2)$-dimensions respectively such that

$$L(\delta_i, \mathsf{Y}_j; \mathsf{V}) = \delta_{ij}/m_j \qquad (1 \leqq i, j \leqq t),$$

where m_j means the common order of δ_j and Y_j for $1 \leqq j \leqq t$. Moreover by the Lefschetz-Hodge theorem in (H), § 51. 2, we may assume that Y_i are all (algebraic) V-divisors. Since Y_i is homologous to 0 modulo Q, there exists a multiplicative function F_i on V such that $(F_i) = \mathsf{Y}_i$ for $1 \leqq i \leqq t$; and we have

$$\chi_{\mathsf{Y}_i}(\delta_j) = \exp(2\pi i \delta_{ij}/m_j) \qquad (1 \leqq i, j \leqq t).$$

Therefore if we denote by $\mathfrak{G}_n(\mathsf{V})$ the group of V-divisors, which are homologous to 0 modulo Q, we have the following *duality theorem.*

The factor group $\mathfrak{G}_n(\mathsf{V})/\mathfrak{G}_l(\mathsf{V})$ is dually paired with the homology group $\mathfrak{H}^1(\mathsf{V}, Z)$ by the multiplication rule

$$\left.\begin{array}{l}\mathfrak{G}_n(\mathsf{V}) \ni \mathsf{Y} \\ \mathfrak{H}^1(\mathsf{V}, Z) \ni \Gamma\end{array}\right\} \mathsf{Y} \cdot \Gamma = \chi_{\mathsf{Y}}(\Gamma).$$

It follows that the group $\mathfrak{G}_n(\mathsf{V})/\mathfrak{G}_l(\mathsf{V})$ does not depend on the choice of the projective model V.

Now we shall show that the group $\mathfrak{G}_n(V)$ coincides with the set of V-divisors, which are *numerically equivalent to* 0. A V-divisor Y is defined to be numerically equivalent to 0, if it satisfies $I(Y, \Gamma; V) = 0$ for every algebraic 1-cycle Γ in V. Since the Kronecker index is always defined, the numerical equivalence defines a group of V-divisors. We note also that the homological equivalence modulo Q is broader than the continuous equivalence and is stricter than the numerical equivalence.

Let $\mathfrak{G}(V)$ be the group of all V-divisors, then the factor group $\mathfrak{G}(V)/\mathfrak{G}_n(V)$ has a finite number $\rho(V)$ of independent generators $Z_1, \cdots, Z_\rho$ such that $\rho(V) \leqq R^{2d-2}(V) = R^2(V)$. Moreover by Proposition 3, if W^2 is a generic 2-section of V over a common field of definition of $Z_1, \cdots, Z_\rho$, then $\Gamma_i = Z_i \cdot W^2$ $(1 \leqq i \leqq \rho)$ are independent algebraic 1-cycles in V modulo Q. It can be readily seen by Lefschetz's theorem and by Hodge's extension of Poincaré's theorem in (H), § 50 that we can find a set of algebraic 1-cycles Γ'_j on the surface W, which are homologous to 0 in V modulo Q, such that Γ_i and Γ'_j form a base of algebraic 1-cycles on W modulo Q and such that $I(\Gamma_i, \Gamma'_j; W) = 0$. Then it follows from Severi's theorem that the intersection matrix

$$(I(Z_i, \Gamma_j; V)) = (I(\Gamma_i, \Gamma_j; W))$$

is non-singular. On the other hand any V-divisor Y can be written uniquely in the form

$$Y \sim \sum_{i=1}^{\rho} a_i Z_i \qquad \text{(in } V \text{ mod. } Q)$$

with Z-coefficients a_i $(1 \leqq i \leqq \rho)$. If Y is numerically equivalent to 0, so is $\sum_{i=1}^{\rho} a_i Z_i$, hence by what we have just proved, this must be 0; which proves the assertion.

The integer $\rho(V)$ is called the *Picard number of* V^d; we shall call the difference $\tau(V) = R^{2d-2}(V) - \rho(V)$ the *Lefschetz number* of V, and we shall prove its absolute invariance. Let $Z_1, Z_2, \cdots, Z_\tau$ be a base of the "transcendental" $(2d-2)$-cycles in V, then $\Gamma_i = Z_i \cdot W^2$ $(1 \leqq i \leqq \tau)$ form a base of the transcendental 2-cycles in V. Therefore by a similar argument as in (H), § 51. 2, we have, for every algebraic differential of degree 2 of the first kind Φ in V,

$$\int_{\sum_{i=1}^{\tau} a_i \Gamma_i} \Phi = 0,$$

with a_i in Z, if and only if $a_i = 0$ $(1 \leqq i \leqq \tau)$. On the other hand let T be a birational correspondence between V and V'. Then we may speak of the image Γ'_i of Γ_i by T for $1 \leqq i \leqq \tau$; and we have

$$\int_{\sum\limits_{i=1}^{\tau} a_i\Gamma'_i} \Phi' = 0$$

for every algebraic differential of degree 2 of the first kind Φ' in V', if and only if $a_i = 0$ $(1 \leqq i \leqq \tau)$. Therefore $\Gamma'_1, \cdots, \Gamma'_\tau$ are independent transcendental 2-cycles in V'. Accordingly we must have $\tau(\mathsf{V}) \leqq \tau(\mathsf{V}')$, and similarly $\tau(\mathsf{V}') \leqq \tau(\mathsf{V})$, which completes the proof.

Appendix II. Algebraic Correspondence and Poincaré Family.

Let $\mathsf{V}_1^{d_1}$ and $\mathsf{V}_2^{d_2}$ be two non-singular projective models with irregularities q_1 and q_2 respectively, and consider the module of $(\mathsf{V}_1 \times \mathsf{V}_2)$-divisors over Z; its member may be called a *correspondence between* V_1 *and* V_2. It can be readily seen that a correspondence X satisfies $\mathsf{X}(z_1) \equiv \mathsf{X}(z_2)$ (mod. $\mathfrak{G}_l(\mathsf{V}_2)$), for every pair of points z_1 and z_2 on V_1, if and only if X is of the form

$$\mathsf{X} \equiv \mathsf{V}_1 \times \mathsf{Y}_2 + \mathsf{Y}_1 \times \mathsf{V}_2 \quad (\text{mod. } \mathfrak{G}_l(\mathsf{V}_1 \times \mathsf{V}_2)),$$

with some V_α-divisors Y_α $(\alpha = 1, 2)$. We shall call such an X a *correspondence with valence* 0. Since the correspondences with valence 0 form a submodule of the module of all correspondences, we can consider the factor module $\mathscr{B}(\mathsf{V}_1, \mathsf{V}_2)$. On the other hand since V_α are orientable manifolds, $\mathfrak{H}^{2d_\alpha-1}(\mathsf{V}_\alpha, Z)$ are free abelian groups with $2q_\alpha$ generators $\Gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$; and if we put $\Gamma_{\alpha i} \cdot \mathsf{W}_\alpha{}^1 = \gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$, where $\mathsf{W}_\alpha{}^1$ are generic 1-sections of V_α, then the $\gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$ form a base of $\mathfrak{H}^1(\mathsf{V}_\alpha, Q)$ for $\alpha = 1, 2$. Moreover every $2(d_1 + d_2 - 1)$-cycle X of $\mathsf{V}_1 \times \mathsf{V}_2$ over Z can be written uniquely in the form

$$X \sim \mathsf{V}_1 \times \Gamma_2 + \Gamma_1 \times \mathsf{V}_2 + \sum_{ij} s_{ij}(\Gamma_{1i} \times \Gamma_{2j}) \ (\text{mod. } Z),$$

where Γ_α are Z-cycles of $2(d_\alpha - 1)$ dimension in V_α $(\alpha = 1, 2)$ and where $S = (s_{ij})$ is a Z-matrix of type $(2q_1, 2q_2)$. Let ω_α be the period matrices of the Picard integrals of the first kind on V_α along the 1-cycles $\gamma_{\alpha i}$ $(1 \leqq i \leqq 2q_\alpha)$ and let A_α be the abelian varieties attached to these Riemann matrices for $\alpha = 1, 2$. It can be readily calculated without difficulties by the Lefschetz-Hodge theorem that X is algebraic if and only if Γ_α are algebraic and S satisfies the relation $\omega_1 S^t \omega_2 = 0$. Therefore if we denote by $\lambda(\mathsf{V}_1, \mathsf{V}_2)$ the rank of the Z-module of such matrices S, we have the following formula, which is well known if V_1 and V_2 are curves:

$$\rho(V_1 \times V_2) = \rho(V_1) + \rho(V_2) + \lambda(V_1, V_2).$$

Now the module $\mathfrak{H}(A_1, A_2)$ of all homomorphisms of A_1 into A_2 can be represented faithfully as the module of "complex multiplications" of ω_1 to ω_2; and if we extend this module over Q, it is isomorphics with the module of all Q-matrices satisfying $\omega_1 S^t \omega_2 = 0$. Moreover if a correspondence X is homologous to 0 modulo Z, we can conclude from Theorem 5 that X is of valence 0. Therefore the Z-modules $\mathfrak{B}(V_1, V)$ and $\mathfrak{H}(A_1, A_2)$ have the same rank; since they are both free, however, we get the isomorphism

$$\mathfrak{B}(V_1, V_2) \cong \mathfrak{H}(A_1, A_2).$$

As a free abelian group, $\mathfrak{H}(A_1, A_2)$ depends only on the categories of A_1 and A_2.

If we apply the above result to the case where $V_1 = A$ is the Albanese variety attached to a given variety $V_2 = V$. we see that $\mathfrak{B}(A, V)$ is generated by the Poincaré family in the "general case" $\mathfrak{H}(A, A) = Z$; and then, up to a scalar factor, the Z-matrix E in our theory is an absolute invariant of V. In this connection it is also to be remarked that there exists a Riemann matrix with any preassigned principal matrix; thus the Albanese variety and the Picard variety attached to the same variety are not isomorphic in general.

KYOTO UNIVERSITY, JAPAN.

CONTOUR EQUIVALENT PSEUDOHARMONIC FUNCTIONS AND PSEUDOCONJUGATES.*

By James A. Jenkins and Marston Morse.

1. Introduction. The present paper deals with the structure of the level curves of pseudoharmonic functions defined in the closure of Jordan domains D and satisfying certain boundary conditions. We are in particular interested in formulating necessary and sufficient conditions that two such functions U and V be *contour equivalent* (written C. E.) that is, that there exist a sense-preserving homeomorphism (written S. P. homeomorphism) ϕ of the domain of U onto that of V under which the connected level arcs of U are mapped onto the connected level arcs of V. By defining and constructing a pseudoconjugate v of an admissible u an interior transformation $u + iv$ of D is obtained. It is proved that for a given admissible U a S. P. homeomorphism ϕ of $\bar{D}$ onto itself exists such that the composite function $U\phi$ is harmonic. The theorems on contour equivalence are new even when U and V are both harmonic.

The set of level curves of a pseudoharmonic function U which emanate from the multiple points of U is termed the *net* of U. We show that each admissible U is C. E. to a model U_0 with a net composed of hyperbolic lines (regarding the disc $|z| \leq 1$ as a hyperbolic plane).

Reference will frequently be made to the book: Marston Morse, "Topological Methods in the Theory of Functions of a Complex Variable," Princeton University Press, which we denote by M.

When a pseudoharmonic function U is defined over a general simply connected *open* domain a pseudoconjugate of U still exists although radically different methods of proof are required. It is even possible to start with families of curves which have the topological properties of level arcs but which are not given as level arcs of a function U. Contact is thus made with some of the results of W. Kaplan [1] although the methods used are quite distinct. Results in this *open* case will be presented later.

In the bibliography, reference is made to two additional papers, of interest in connection with this paper.

* Received February 12, 1951.

[1] W. Kaplan, "Topology of level curves of harmonic functions," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 514-522.

Part I. Contour Equivalence.

2. Fundamental definitions. A function U with real values $U(z)$, $z = x + iy$, is said to be pseudoharmonic at the point z_0 of the z-plane if $U(z)$ is defined in a neighborhood N of z_0 and if there exists a S. P. homeomorphism ϕ from a neighborhood N_1 of z_0 to N, leaving z_0 fixed, such that the composite function $U\phi$ is harmonic and non-constant on N_1. A critical point of $U\phi$ is termed a *critical* point of U. A function U is said to be pseudoharmonic on a domain D (open) if it is pseudoharmonic at each point of D. A subset of D on which $U(z)$ is contant will be called a *U-set.* A *U-arc* is thus well-defined. We understand that a U-set but not a U-arc may reduce to a point. We shall introduce canonical neighborhoods N of a point $z_0 \varepsilon N$ and canonical representations of U over N as follows.

Case I. Let U be pseudoharmonic in a neighborhood of z_0. There then exists a neighborhood N of z_0 free from critical points of U except at most z_0, with $\bar{N}$ the homeomorph of a plane circular disc such that z_0 corresponds to the center of the disc and the locus on which $U(z) = U(z_0)$ in N corresponds to a set of $2n$ rays $(n > 0)$ leading from this center and making successive sectors of central angle π/n. As a variable point z crosses any one of these U-arcs (except at z_0) the difference $U(z) - U(z_0)$ changes sign. (See M., Th. 2.1.) If $n = 1$, a single U-arc passes through z_0 and z_0 is termed *ordinary.* If $n > 1$ z_0 is termed a *multiple* point of U of *index* $2n - 2$. A neighborhood of z_0 such as N will be termed *canonical.*

Let D be a Jordan domain bounded [2] by a Jordan curve βD.

CONDITIONS [3] $[D]$. *A real valued function U defined over $\bar{D}$ will be said to be in $[D]$ if U is pseudoharmonic on D, continuous on $\bar{D}$, and if $U \mid \beta D$ assumes its relative extreme values in at most a finite set of points in βD.*

We recall a number of properties of a $U \varepsilon [D]$. Every point $z_0 \varepsilon \beta D$ has a canonical neighborhood coming under Case II or Case III.

Case II. $z_0 \varepsilon \beta D$; z_0 *not a point of relative extremum of U.* Here z_0 has a canonical vicinity N relative to $\bar{D}$, free from critical points (cf. M) of $\pm U$ in $\bar{D}$, with $\bar{N}$ the homeomorph of a semi-disc H, such that z_0 corresponds to the center O of H, $\bar{N} \cap \beta D$ corresponds to the diameter of H, while the U-set in N at the level $c = U(z_0)$ is represented by n rays $(n > 0)$ emanating

---

[2] As a general notation, βD shall mean the boundary of D.

[3] Conditions $[D]$ include conditions called Boundary Conditions A in Morse, loc. cit.

from O, and dividing N into $m = n + 1$ sectors (open) on which $U - c$ alternates in sign. If $m > 2$, z_0 is termed a boundary *multiple* point of U of *index* $m - 1$.

Case III. $z_0 \varepsilon \beta D$; *z_0 a point of relative extremum of U.* In this case there exists a canonical neighborhood N of z_0 relative to $\bar{D}$ free from critical points of $\pm U$ with $\bar{N}$ the homeomorph of a semi-disc such that z_0 corresponds to the center of the semi-disc, $\bar{N} \cap \beta D$ corresponds to the diameter of the semi-disc, the boundary of N in D is at a level $c_1 \neq c$ such that for $z \varepsilon N$ and $z \neq z_0$, $U(z)$ is between c_1 and c. (M., Th. 7.2.) The point z_0 is termed *ordinary* in this case.

U-continuation. As shown in M., § 7, a $U \varepsilon [D]$ has only a finite number of multiple points in $\bar{D}$. By a U-continuation of a simple U-arc g will be meant a simple U-arc containing g which continues through a multiple point z_0 of U in D with the U-arc which issues from z_0 opposite to g in the canonical representation of U neighboring z_0.

CONTOUR EQUIVALENCE. *Let Δ be a second Jordan domain. Two functions U and V continuous over $\bar{D}$ and $\bar{\Delta}$ respectively, are said to be* C. E. *under ϕ, if there exists a* S. P. *homeomorphism ϕ of $\bar{D}$ onto $\bar{\Delta}$ such that each maximal connected U-set (V-set) corresponds to a like V-set (U-set); ϕ is said to define a strict contour equivalence if $U = V\phi$.*

Examples show that not every contour equivalence of pseudoharmonic functions is strict. However, we shall prove in § 7 that a contour equivalence of a $U \varepsilon [D]$ with a $V \varepsilon [D]$ implies a strict contour equivalence, if one admits a preliminary transformation $U' = -U$ (if necessary) and a preliminary continuous deformation of U through functions in $[D]$ with fixed level arcs. One sees that contour equivalence, as well as strict contour equivalence, is symmetric, reflexive, and transitive.

Type of multiple point. In § 4 we shall enlarge the class $[D]$ to the class $[D]'$. Understanding the term multiple point P and canonical neighborhood $N(P)$ in the enlarged sense of § 4 as well as in the sense of § 2, let P and Q be multiple points of U with canonical neighborhoods $N(P)$ and $N(Q)$. We term P and Q multiple points of the *same* type if for some choice of $N(P)$ and $N(Q)$, $\bar{N}(P)$ admits a homeomorphism onto $\bar{N}(Q)$ which maps U-arcs onto V-arcs.

It is readily seen that multiple points of the same type have equal indices and both come under the same one of the Cases I, II (or V, VII in the enlarged sense).

Interior transformations f. A function f with complex values $f(z)$ in the $w = u + iv$-plane will be said to be *interior at z_0* if f is defined in a neighborhood N of z_0, and if there exists a S. P. homeomorphism ϕ of a neighborhood N_1 of z_0 onto N leaving z_0 fixed and such that the composite function $f\phi$ is meromorphic and non-constant on N_1. We term f *interior* over D if f is interior at each point $z_0 \varepsilon D$. If f is interior over D the real and imaginary parts of f are pseudoharmonic over D apart from poles of f.

The following two lemmas will be useful in § 11, § 12.

LEMMA 2. 1. *Let z_0 be a point in D. If f is interior in $D - z_0$ and continuous in D, then f is interior at z_0.*

Set $f(z_0) = w_0$. For $e > 0$ sufficiently small, the set $E_e: | z - z_0 | \leqq e$ is in D and for some such e the image curve $g = f(\beta E_e)$ does not intersect w_0. Let e be so chosen. Let m be the order of g with respect to w_0. It is seen that each point $w \neq w_0$ in a sufficiently small neighborhood of w_0 is covered m times by $f \,|\, (E_e - z_0)$, so that f is open and light. The lemma follows.

LEMMA 2. 2. *If f is interior in D, continuous in $\bar{D}$ and if $f(\beta D) = g$ is locally simple, then f admits an interior extension over some neighborhood of any given point z_0 of βD.*

As seen in M., p. 85 the points $z_0 \varepsilon \beta D$ neighboring which f fails to be topological are isolated in βD. Lemma 2. 2 is immediate except at a point z_0 neighboring which f fails to be topological. Suppose then that z_0 is such a point.

As seen in M., § 23 one can suppose without loss of generality that $z_0 = 0$ and that there is an open arc h of βD of the form $[y = 0, -e < x < e]$ with an image $f(h)$ covering an arc $[v = 0, a < u < b]$ in a 1-1 manner. We continue f over h by requiring that $f(\bar{z}) = \bar{f}(z)$ for points $z \varepsilon \bar{D}$ near $z = 0$. It is seen that the extended f satisfies the conditions of Lemma 2. 1 in a sufficiently small neighborhood of $z_0 = 0$ and so is interior at z_0.

3. Conditions necessary for contour equivalence. Contour equivalence of $U \varepsilon [D]$ with $V \varepsilon [\Delta]$ under ϕ implies a 1-1 mapping T of the set of multiple points of U onto the set of multiple points of V, sending a multiple point of U into a multiple point of V of the same type, and boundary multiple points into boundary multiple points in the same cyclic order. If a finite 1-1 correspondence T of this type is given without giving ϕ but satisfying certain other finite conditions to be enumerated, there then exists (§ 6) an extension ϕ of T over $\bar{D}$ such that ϕ defines a contour equivalence of U with V. New terms are needed.

The star $S(P, U)$. Let P be a multiple point of U in D or βD. Let each U-arc g with end point at P be "continued" (§ 2) on D (open) from P until a point of βD is reached. So continued and sensed g will be termed a *ray* of the *star* $S(P, U)$ of P. The rays of $S(P, U)$ will be simple and non-intersecting except at P. They will divide D into open connected sectors

$$S_1(P, U), \cdots, S_m(P, U) \qquad (m > 2) \tag{3.1}$$

where m is even when $P \varepsilon D$. We suppose that these sectors are indexed in the order in which their boundary arcs on βD follow each other in counter-clockwise sense, starting with the first such arc of βD following P in case $P \varepsilon \beta D$. For $P \varepsilon \beta D$, $S_1(P, U)$ will then be uniquely determined. For $P \varepsilon D$, $S_1(P, U)$ can be taken as any one of the sectors of $S(P, U)$. The boundary of $S_i(P, U)$ will include an arc of βD and one or two rays of $S(P, U)$. We term P the *center* of the star $S(P, U)$. Two stars will be said to be of the same *type* if their centers are of the same type.

Frames $M(U)$ and mappings T. We shall obtain homeomorphisms ϕ defining contour equivalence of pseudoharmonic functions U and V as extensions over $\bar{D}$ of mappings T of special subsets (§ 4) $M(U) \subset \bar{D}$. To be admissible a *frame* $M(U)$ shall include all the multiple points of U. A mapping T of $M(U) \subset \bar{D}$ onto $M(V) \subset \bar{\Delta}$ to be admissible must be a homeomorphism which preserves the type of each multiple point and which maps $M(U) \mid \beta D$ onto $M(V) \mid \beta\Delta$ with preservation of sense in case $M(U) \mid \beta D$ contains at least three points. With this understood we define similarity of $M(U)$ with $M(V)$ under T relative to two stars $S(P, U)$ and $S(Q, V)$ for which $Q = TP$.

RELATIVE SIMILARITY UNDER T. *Let T be an admissible mapping of $M(U)$ onto $M(V)$ with a multiple point P of U going into a multiple point Q of V of the same type. We say that $[M(U), S(P, U)]$ is similar to $[M(V), S(Q, V)]$ under T if for some admissible indexing of $S(P, U)$ and $S(Q, V)$ and for each point $z \varepsilon M(U)$ and image $w \varepsilon M(V)$ under T the incidence relations*

$$z \varepsilon \bar{S}_i(P, U), \qquad w \varepsilon \bar{S}_i(Q, V) \qquad (i = 1, \cdots, n) \tag{3.2}$$

both hold or both fail to hold for each i.

SIMILARITY UNDER T. *Admissible sets $M(U) \subset \bar{D}$ and $M(V) \subset \bar{\Delta}$, homeomorphic under an admissible mapping T, are termed similar under T if for any two multiple points P and Q which are images under T, $[M(U), S(P, U)]$ is similar to $[M(V), S(Q, V)]$ under T.*

SIMILARITY. *Admissible sets $M(U) \subset \bar{D}$ and $M(V) \subset \bar{\Delta}$ are termed similar (written $M(U) \sim M(V)$) if $M(U)$ and $M(V)$ are similar under some admissible T.*

It is clear that the relation of similarity between admissible frames $M(U)$ and $M(V)$ is reflexive, symmetric, and transitive. It is thereby not excluded that $U = V$ and that $M(U) \sim M(U)$ under some T other than the identity.

The basic conditions for contour equivalence are given in Th. 3.1. That these conditions are sufficient will be proved in § 6.

THEOREM 3.1. *In order that $U \varepsilon [D]$ and $V \varepsilon [\Delta]$ be* C. E. *under ϕ it is necessary and sufficient that the set $\mu(U)$ of multiple points of U and the set $\mu(V)$ of multiple points of V be similar under some admissible mapping T.*

The condition is necessary, since the existence of the mapping ϕ implies the existence of the mapping $T = \phi \mid \mu(U)$ of $\mu(U)$ onto $\mu(V)$. Moreover, T is admissible in that T, like the mapping ϕ, preserves the type of the multiple point, carries boundary multiple points into boundary multiple points and preserves cyclic order of these boundary points.

In satisfying the condition of similarity of $\mu(U)$ with $\mu(V)$ an indexing (3.1) of the sectors of $S(P, U)$ must be made and coordinated with an indexing of the sectors of $S(Q, V)$ where $T(P) = Q$ so that the incidence conditions (3.2) both hold or fail to hold. When P and Q are boundary points only one indexing is possible. If P and Q are in D and happen to be the only multiple points of U and V respectively, Th. 3.1 is vacuous except for the condition that P and Q have the same multiplicity; in this case the sectors $S_1(P, U)$ and $S_1(Q, V)$ can be chosen arbitrarily from among the sectors of $S(P, U)$ and $S(Q, V)$ respectively.

If U possesses a multiple point P' different from P, and if V is C. E. with U, then V possesses the multiple point $T(P') = Q' \neq Q$. Given an indexing (3.1) of the sectors of $S(P, U)$, an indexing of the sectors of $S(Q, V)$ is then uniquely determined by the condition that the sector (or two adjacent sectors) of $S(P, U)$ with whose closure (closures) P' is incident, bear the same index (indices) as the sector (or two adjacent sectors) of $S(Q, V)$ with whose closure (closures) Q' is incident.

4. Conditions $[D]'$. To establish the sufficient conditions for contour equivalence of U and V as given in Th. 3.1 it is convenient to establish a similar theorem in which the preceding class $[D]$ is enlarged. In § 12 we shall show that a $U \varepsilon [D]$ admits a pseudoharmonic continuation over an open domain $D' \supset \bar{D}$. If a U were constant on some arc of βD but otherwise satisfied the conditions $[D]$ this pseudoharmonic continuation over βD might prove impossible, as examples [4] would show. This fact motivates the following definition.

CONDITIONS $[D']$. *A function U defined and continuous at each point of $\bar{D}$ will be said to be in $[D]'$, if U is pseudoharmonic over D, if there are at most a finite set of maximal connected U-arcs in βD, if $U \mid \beta D$ assumes its relative extrema in at most a finite set of such U-arcs or points in βD, and if U admits a pseudoharmonic continuation over βD.*

Suppose $U \varepsilon [D]'$. If z_0 is in a maximal connected U-arc g in βD, a canonical neighborhood N of z_0 and local representation of U over N will not come under Cases I, II, III of § 2. The local representations described below are a ready consequence of the assumption that U admits a pseudoharmonic continuation over βD. There are four new cases.

IV. z_0 interior to g, and a point of relative extremum [5] of U.

V. z_0 interior to g and *not* a point of relative extremum of U.

VI. z_0 an end point of g, and a point of relative extremum of U.

VII. z_0 an end point of g, and *not* a point of relative extremum of U.

In a canonical neighborhood N (relative to $\bar{D}$) of a point z_0 a *sector* of N shall be understood as any maximal connected subset of N on which $U(z) \neq U(z_0)$.

Case IV. A canonical N exists with $\bar{N}$ the homeomorph of a semi-disc H such that z_0 corresponds to the center O of H, $\beta N \cap \beta D$ corresponds to the diameter d of H, and maximal U-arcs in N correspond to chords in H parallel to d. Cf. Case III, § 2.

Case V. An N exists essentially as in Case II except that *all* arcs

[4] One could construct an example in which a U-arc in D had every point of a U-arc in βD as limit point.

[5] That is, in some neighborhood of z_0 relative to $\bar{D}$ $U(z) \geqq U(z_0)$, or else $U(z) \leqq U(z_0)$.

bounding the sectors of N and emanating from z_0 are U-arcs (i. e., including the two arcs of βD emanating for z_0).

Case VI. An N exists homeomorphic with a semi-disc H, such that z_0 corresponds to the center O of H, and $\beta N \cap \beta D$ corresponds to the diameter d of H. Of the two rays ρ_1, ρ_2 of d separated by O, U has its extreme value c at each point of one, say ρ_1, and is strictly monotone on the other. An arbitrary maximal U-arc in N not at the level c corresponds to a quarter circle which starts at a point of ρ_2, has O as its center, and is continued by a straight arc in H parallel to d until βH is reached.

Case VII. An N exists similar to the N under Type V, except that *one* of the sectors of N (the initial or final sector) has one boundary arc in βD, not a U-arc.

Proofs of these statements can be made with complete rigor, using the representations in Cases I, II, and III of the pseudoharmonic extension of U over βD.

Theorem 4.1 depends for its meaning on the definition of an admissible mapping T of § 3, and this in turn depends upon the meaning attached to two multiple points being of the *same type*. The necessary definitions will now be given.

A canonical neighborhood N of a point $z_0 \varepsilon \beta D$ contains just one sector in Cases III, IV, and VI, and in each of these cases is termed *ordinary*. Let m be the number of sectors in a canonical neighborhood of z_0. In Cases V or VII, $m > 1$ and z_0 is termed a boundary *multiple* point of index $m - 1$. It is seen that a point z_0 of intersection of a U-arc in βD with any U-arc not in βD (except for z_0) always comes under Case V or VII, and so is always a multiple point of positive index.

Given $U \varepsilon [D]'$ and $V \varepsilon [\Delta]'$ let P and Q be multiple points of U and V respectively. For P and Q to be multiple points of the *same type* (as defined in § 3) it is necessary and sufficient that P and Q come under the same one of the Cases I, II, V, or VII and have equal positive indices.

As in § 3 an admissible *frame* $M(U)$ must contain each multiple point of U. Stars $S(P, U)$ are formally defined as in § 3 admitting the new types of multiple points P. As previously two stars $S(P, U)$ and $S(Q, V)$ are termed of the *same type* if P and Q are of the same type. Recall that an admissible mapping T of a frame $M(U)$ onto a frame $M(V)$ is a homeomorphism mapping the set of multiple points of U onto the set of multiple points of V, preserving the type of a multiple point and mapping $M(U) \mid \beta D$ onto $M(V) \mid \beta\Delta$ with preservation of sense on the respective boundaries.

The conditions necessary and sufficient for contour equivalence of U and V include those of Th. 3.1 and are given as follows.

THEOREM 4.1. *Given $U \varepsilon [D]'$ and $V \varepsilon [\Delta]'$ let $M(U)[M(V)]$ be the union of the multiple points of $U[V]$ with U-arcs [V-arcs] in βD, $[\beta\Delta]$. In order that U and V be* C. E. *under a* S. P. *homeomorphism ϕ it is necessary and sufficient that $M(U)$ be similar to $M(V)$ under an admissible mapping T of $M(U)$ onto $M(V)$. Given T, ϕ can be taken as an extension of T over $\bar{D}$. Given ϕ, T can be taken as $\phi \mid M(U)$.*

That the conditions of the theorem are necessary is an immediate consequence of the definitions. That the conditions are sufficient will be established in the next sections. In the next section certain intuitive notions needed in the proofs are made precise.

5. Local right sets, and sensing of U-arcs. Let g be a sensed Jordan arc and $z = a$ an inner point of g. We shall make the intuitive notion of the right of g near a more precise. To that end let N be a Jordan region containing a whose intersection with g is a subarc g' of g forming a cross cut of N. There then exists a S. P. homeomorphism H of N onto the disc $(|z| < 1)$ such that g' is mapped onto the positively sensed segment of the real axis in the disc. The inverse image in N of the set $(|z| < 1, y < 0)$ [alternatively $|z| < 1, y > 0$,] will be called a *local right* [*left*] set of the element (a, g). It is readily shown that for a given element (a, g) local right sets and local left sets of sufficiently small diameter have an empty intersection, while the intersection of any two right (left) sets includes a right (left) set. S. P. homeomorphisms which (by convention) preserve the sense of an arc g, carry right (left) sets of an element (a, g) into right (left) sets of the image element.

Let $z = a$ be an ordinary point of a pseudoharmonic function U. There exists a S. P. homeomorphism ϕ of a neighborhood of $z = a$ into the w-plane $(w = u + iv)$ carrying $z = a$ into $w = 0$ and U-arcs into curves on which v is constant. If g is a properly sensed U-arc through $z = a$, a right set (left set) of (a, g) of sufficiently small diameter thus consists of the points z in a neighborhood of a such that $U(z) < U(a)$ $(U(z) > U(a))$. Such a sensing of g will be termed *U-positive,* the opposite sensing *U-negative.* Such a U-positive sensing of a U-arc g near an ordinary point $z = a$ in g is independent of the mappings and neighborhoods used to define their sensing. If U is pseudoharmonic in D it is possible to assign a positive sense to each U-arc composed of ordinary points of U such that this assignment agrees with each local assignment of a U-positive sense to a U-arc. At multiple points of U, this sensing is ambiguous.

A particular application of this is as follows. Let a $U \varepsilon [D]'$ assume its absolute minimum on an arc p of βD. Then a U-positive sensing of p will be counter-clockwise on βD. If however U assumes its absolute maximum on an arc q of βD, a U-positive sensing of q will be clockwise.

6. Sufficient conditions for contour equivalence. We shall prove that the conditions of Th. 4.1 are sufficient. Let $\nu(U)$ be the sum of the indices of the multiple points of U, and let $\nu(V)$ be similarly defined. We shall make an induction depending upon the value of $\nu(U) = \nu(V)$ recalling that $\nu(U) = \nu(V)$ when the conditions of Th. 4.1 are satisfied. It is necessary to set $\nu(U) = 0$ if U has no multiple points, similarly for $\nu(V)$. By convention let each ordinary point have an index 0.

LEMMA 6.1. *The truth of* Th. 4.1 *for* $\nu(U) = \nu(V) < n$, *and* $n > 0$ *implies its truth for* $\nu(U) = \nu(V) = n$.

Arcs p and q. It is given that the frames $M(U)$ and $M(V)$ appearing in Th. 4.1 are similar under a homeomorphism T. Let P_0 and Q_0 be fixed multiple points of U and V respectively, corresponding under T. The conditions for star similarity under T of $[M(U), S(P_0, U)]$ and $[M(V), S(Q_0, V)]$ as defined in § 3 are satisfied (by hypothesis) after a suitable indexing of the sectors of $S(P_0, U)$ and a corresponding indexing of the sectors of $S(Q_0, V)$. By virtue of this indexing each ray of $S(P_0, U)$ corresponds to a definite ray of $S(Q_0, V)$. Let p be an arbitrary ray of $S(P_0, U)$ in case $P_0 \varepsilon \beta D$, and in case $P_0 \varepsilon D$ let p be the U-continuation in both senses to βD of an arbitrary ray of $S(P, U)$ with p sensed as the ray. Let q be the corresponding ray of $S(Q_0, V)$, or its continuation in case $Q_0 \varepsilon \Delta$.

By virtue of the assumed similarity of $M(U)$ and $M(V)$ under T, and the choice of p and q, the multiple points of U in p correspond in a 1-1 manner under T to the multiple points of V in q with the order in which the multiple points on p appear on p as a sensed arc the same as the order in which their T-images appear on q as a sensed arc. One sees this on considering incidences with the stars $S(P, U)$ and $S(Q, V)$ where P is an arbitrary multiple point on p and $T(P) = Q \varepsilon q$.

A simplifying modification. Without any loss of generality we can suppose that $\bar{D}$ is the disc $(|z| \leqq 1)$; for under a S. P. homeomorphism ϕ of this disc onto $\bar{D}$, U is C. E. with $U\phi$ (defined over the disc).

The proof can accordingly be simplified by supposing that $\bar{D}$ and $\bar{\Delta}$ are each the disc $(|z| \leqq 1)$, that p and q are each the diameter d of this disc leading from $z = -1$ to $z = +1$, and that the multiple points z^* of U in d are identical with those of V in d, with $T(z^*) = z^*$. The point $z = 1$, if in $M(U)$, is a multiple point of U of Case II, V or VII; likewise the point $z = -1$. Granting this a priori simplification, we assume as previously that $M(U) \sim M(V)$ (under T).

Let D_1 be the open upper semi-disc of D: $(|z| < 1)$. Set $u = U \mid \bar{D}_1$, $v = V \mid \bar{D}_1$. Let $M(u)[M(v)]$ be the union of the multiple points of u, $[v]$ with the u-arcs [v-arcs] in βD_1.

The mapping T_1. With u and $v \,\varepsilon\, [D_1]'$ an admissible mapping T_1 (see § 3) of $M(u) \subset \bar{D}_1$ onto $M(v) \subset \bar{D}_1$ is obtained on setting $T_1(z) = T(z)$ for $z \,\varepsilon\, M(U) \cap \bar{D}_1$ and $T_1(z) = z$ for $z \,\varepsilon\, d$. These two conditions on $T_1(z)$ both apply to a point z of $M(U) \cap d$, and are consistent. This is true if $z \,\varepsilon\, d$ is a multiple point of U, since $T(z) = z$, as arranged in the preceding paragraphs. It is true if $z = \pm 1$ is in $M(U)$, since this can happen only if $z = \pm 1$ is a multiple point of U [Cf. Cases V and VII of § 4.] The mapping T_1 of the frame $M(u)$ onto the frame $M(v)$ is admissible in the sense of § 3 in that it maps the set of multiple points of u onto the set of multiple points of v, preserving type, and maps $M(u) \mid \beta D_1$ onto $M(v) \mid \beta D_1$ preserving sense on βD_1.

The indexing of stars $S(P, u)$ and $S(Q, v)$. If P and Q are multiple points of u and v respectively with $T_1(P) = Q$ then P and Q are multiple points of U and V with $T(P) = Q$. The sectors of $S(P, u)$ and $S(Q, v)$ will be indexed in such a manner that a sector $S_i(P, u)$ has the index i in common with a sector $S_i(Q, v)$ if and only if the sectors in $S(P, U)$ and $S(Q, V)$ containing $S_i(P, u)$ and $S_i(Q, v)$ respectively as point sets, bear equal indices k.

Finally $M(u) \sim M(v)$ under T_1. The analysis follows. Each multiple point of u or v respectively is a multiple point of U or V. Among points in $\bar{D}_1$, the converse is true except at most for the points $z = \pm 1$. The stars of u [or v] are thus the intersections with $\bar{D}_1$ of stars of U [or V] with the same centers. Note that $M(u) = [M(u) \cap M(U)] \cup d$. Hence $M(u) = [M(u) \cap M(U)] \cup A$, where A is the set of points $z_0 \,\varepsilon\, d$ not multiple points of U.

The incidence relations. It follows from the definition of T_1 and the indexing of the sectors of the stars $S(P, u)$ and $S(Q, v)$ that a point

$z^* \varepsilon M(u) \cap M(U)$ has the same incidence relations with closed sectors of $S(P, u)$ as $T(z^*)$ with the corresponding closed sectors of $S(Q, v)$. These points z^* include the multiple points of u in d (in particular possibly $P_0 = Q_0$). These multiple points z^* of u in d are also multiple points of U, V and v with $T(z^*) = T_1(z^*) = z^*$.

It remains to consider a point $z_0 \varepsilon A$. Since z_0 is not a multiple point of U no ray of a star of U or V meets z_0 other than a ray on d. The point z_0 is immediately preceded (or followed) on d by a multiple point z^* of U, V. If z^* is on a sector $S_i(P, u)$ with P not in d, or on a left (or right) boundary ray of $S_i(P, u)$ then z_0 is in $S_i(P, u)$. It follows that z^* is in $S_i(Q, v)$ or on the left (or right) boundary ray of $S_i(Q, v)$. Hence z_0 is in $S_i(Q, v)$. Thus z_0 has the same incidence relations with the closed sectors of $S(P, u)$ as with the corresponding closed sectors of $S(Q, v)$. If $P = Q$ is in d with x-coordinate $a < 1$ then the subarc of d on which $a \leqq x \leqq 1$ is a right boundary arc of the first sector both of $S(P, u)$ and $S(Q, v)$ so that any point x_0 on this arc has the same incidence relations with the closed sectors of $S(P, u)$ as with the closed sectors of $S(Q, v)$. The arc $-1 \leqq x \leqq a$ is similarly treated if $a > -1$. Thus $M(u) \sim M(v)$ under T_1.

The induction. Suppose that $\nu(U) = \nu(V) = n > 0$. Let $\nu(u) = \nu(v) = m$. The inductive hypothesis of the lemma can be applied to u and v if $m < n$. That $m < n$ follows from the fact that each multiple point P of u and v is a multiple point of U and V of no less index, while each multiple point of U and V on d, in particular P_0, has an index relative to u and v which is less than its index relative to U and V. Indeed P_0 may be ordinary relative to u and v. By virtue of the inductive hypothesis u is C. E. with v under a mapping ϕ which is an extension over $\bar{D}_1$ of T_1 over $M(u)$. Thus $\phi(z) = z$ for $z \varepsilon d$, and $\phi(z) = T(z)$ for $z \varepsilon M(U) \cap D_1$. Let D_2 be the semi-disc of D on which $y < 0$ and set $u' = U \mid \bar{D}_2$, $v' = V \mid \bar{D}_2$. As just shown for u and v, u' and v' are C. E. under a mapping ϕ' such that $\phi'(z) = z$ for $z \varepsilon d$ and $\phi'(z) = T(z)$ for $z \varepsilon M(U) \cap \bar{D}_2$. Let a mapping Φ of $\bar{D}$ onto itself be defined by combining the mappings ϕ and ϕ'. Then U and V are C. E. under Φ, and Φ is an extension of T.

This completes the proof of the lemma.

The inductive proof of Th. 4. 1 will be completed by proving the following. This is the case $\nu(U) = \nu(V) = 0$.

LEMMA 6. 2. *Suppose that $U \varepsilon [D]'$ and V in $[\Delta]'$ have no multiple points and that T is a topological mapping of the maximal connected U-arcs* *in βD onto the maximal connected V-arcs in $\beta\Delta$ preserving sense on the boundaries.*[6] *There then exists an extension ϕ of T over D under which U is* C. E. *to V.*

We begin by verifying the following.

(α). *The only maximal connected U-arcs in βD are those on which U assumes a proper extremum.*

If h were a maximal U-arc in βD on which U was not a proper relative extremum there would be some U-arc in D which would have an end point $P \varepsilon h$. Such a point P would be a multiple point of U coming under Cases V or VII of § 4, contrary to the assumption that U has no multiple points.

(β). *The extreme values of U reduce to an absolute minimum and absolute maximum assumed respectively in just one maximal connected U-set p and one maximal connected U-set q, where p and q may be U-arcs or points.*

Suppose that there were at least two disjoint maximal connected U-sets affording relative minima. Then for a suitable choice of c the set U_c on which $U(z) \leqq c$ would not be connected. Since the set U_c is connected for $c = \max U(z)$ there exists a superior limit c_0 of the values of c for which U_c is not connected. One proves easily as in M., § 10 that there must be a multiple point P at the U-level c_0. The lemma follows.

Reduced μ-length [M., § 27]. If a and b are constants with $a \neq 0$, $aU + b$ is C. E. with U under the identity. No generality is accordingly lost in proving the lemma if we assume that the range of values of U and V is the interval $[0, 1]$. No maximal level set of U or V will then reduce to a point with the possible exception of sets at the level 0 or 1. Each U-arc λ will be referred to its reduced μ-length ρ as parameter. This is the μ-length of λ measured in λ's U-positive sense from λ's initial point on βD, and divided by the total μ-length of λ. On each such λ, ρ varies from 0 to 1 inclusive. The V-arcs will be similarly referred to their reduced μ-lengths σ as parameter.

To begin the proof proper consider first the case in which p and q are maximal connected U-arcs on which U assumes its absolute minimum and maximum respectively, and let $T(p) = p'$ and $T(q) = q'$. The arcs p' and q' may be maximizing and minimizing respectively rather than minimizing and maximizing. No generality in the proof will be lost if we assume that p' and q'

[6] One gives the boundaries a counter-clockwise sense as a basis of reference.

are respectively minimizing and maximizing since a change from V to the C. E. — V would bring this about in any case.

Suppose that the given mapping T of p onto p' has the form $\sigma = \eta(\rho)$, $0 \leqq \rho \leqq 1$, making the point on p with reduced μ-length ρ correspond under T to the point on p' with reduced μ-length σ. The senses of increasing μ-length on p and p' have been taken as the U-positive and V-positive senses on p and p' and so are counter-clockwise on βD and $\beta\Delta$ respectively. From this and the nature of T as given, it follows that $\eta(\rho)$ is *increasing*. Similarly suppose that the given mapping T of q and q' has the form $\sigma = \zeta(\rho)$, $0 \leqq \rho \leqq 1$ and verify the essential fact that $\zeta(\rho)$ is increasing. An admissible extension ϕ of T mapping $\bar{D}$ onto $\bar{\Delta}$ is obtained by making each U-arc λ_c at the level c correspond to the V-arc θ_c at this level c, and making the point ρ on λ_c correspond to the point σ on θ_c such that

$$\sigma = (1 - c)\zeta(\rho) + c\eta(\rho) \qquad (0 \leqq \rho \leqq 1). \tag{6.1}$$

Recalling that $\zeta(\rho)$ and $\eta(\rho)$ are both increasing it follows from the properties of reduced μ-length that ϕ extends T as a sense preserving homeomorphism and defines a contour equivalence of U with V. One first verifies that T is a homeomorphism. That T is sense preserving then follows from the fact that it is sense preserving in the neighborhood of one point, in particular in the neighborhood of a point on p.

In case p reduces to a point but q does not, let $T(q) = q'$, and as before suppose V assumes its absolute maximum on q'. Then the maximal connected V-set in which V assumes its absolute minimum must reduce to a point which we denote by p'. Now define reduced μ-lengths on the level arcs and as in the preceding paragraph let the given mapping T of q onto q' have the form $\sigma = \zeta(\rho)$. Then the mapping $\sigma = \zeta(\rho)$ $(0 < c \leqq 1)$ yields the extension ϕ of T provided we require $\phi(p) = p'$.

In case both the minimizing set p and maximizing set q for U reduce to points, the maximal connected V-sets p' and q' in which V assumes its absolute minimum and maximum also reduce to points. Defining reduced μ-length as before, the mapping $\sigma = \rho$ for $0 < c < 1$ together with $\phi(p) = p'$, $\phi(q) = q'$ provide a mapping ϕ which is a contour equivalence of U with V.

This completes the proof of Th. 4. 1. Th. 4. 1 implies Th. 3. 1 as a special case.

7. A group of operators. Given $U \varepsilon [D]$ let $[U]$ be the class of functions in $[D]$ C. E. with U. In this section we shall describe a multiplicative group of operations which generates $[U]$ from U. Certain preliminary remarks are needed.

The sum of two pseudoharmonic functions u and v defined over D is not in general pseudoharmonic over D. For example, set $z = x + iy$ and let

$$u(z) = x, \qquad v(z) = \tfrac{1}{4}x^2 - x \qquad (|z| < 1).$$

The functions u and v have no critical point on the domain $|z| < 1$ and so are pseudoharmonic. However $u + v$ has a minimum when $z = 0$ and so is not pseudoharmonic. The functions u and v in this example are easily seen to be C. E. under the identity. Thus the sum of two functions C. E. under the identity need not be pseudoharmonic. There is nevertheless a law in the background.

Suppose that u and v are in $[U]$ and C. E. under ϕ. Let g be any simple sensed arc in $\bar{D}$ on which u is strictly increasing. Then for each choice of such a g, v will be strictly increasing on $\phi(g)$, or else strictly decreasing independently of the choice of such a g. This may be verified first for arcs g which intersect the set of multiple points of u in at most an end point. For any one such arc g' can be continuously deformed into any other such arc g'' through admissible arcs g. Finally an admissible arc g which intersects the set of multiple points is a sequence of a finite set of admissible arcs g_i each one of which intersects the set of multiple points in at most an end point.

If g is admissible in the above sense and v is strictly increasing on $\phi(g)$, u and v will be termed *positively* C. E. under ϕ, otherwise *negatively* C. E. under ϕ. Whether u and v are positively or negatively C. E. under ϕ is thus determined by the behavior of v on $\phi(g)$ for one arbitrarily chosen admissible arc g. The following are readily verified:

(α). *If u and v in $[U]$ are* C. E. *under ϕ, then either v or else $-v$ is positively C. E. to u under ϕ.*

(β). *The sum of $u \varepsilon [U]$ and $v \varepsilon [U]$, with u and v given as positively* C. E. *under the identity, is again in $[U]$ and positively* C. E. *to u and to v under the identity.*

To generate $[U]$ from U we shall make use of a deformation Δ of elements $u \varepsilon [U]$. The deformation of u is defined by a 1-parameter family of functions in $[U]$ which for fixed t, $0 \leqq t \leqq 1$, and fixed $u \varepsilon [U]$ have values $\Delta(z, t, u)$ $(z \varepsilon \bar{D})$. For fixed u, Δ is supposed continuous over the cartesian product of the domains of z and t. One supposes that u is given "initially" in the form $u(z) = \Delta(z, 0, u)$. The "terminal" image Δu of u under Δ has by definition the values $(\Delta u)(z) = \Delta(z, 1, u)$ $(z \varepsilon \bar{D})$. For fixed $t \varepsilon [0, 1]$

and $u \varepsilon [U]$, $\Delta(z, t, u)$ shall define a function in $[U]$, C. E. with u under the identity.

With this understood we introduce three operations on elements in $[U]$.

(i) *A homotopy* Δ, *replacing* $u \varepsilon [U]$ *by its terminal image* Δu *under* Δ.

(ii) *A reflection* R, *replacing* u *by* $Ru = -u$.

(iii) *A value-equivalence* Φ, *replacing* u *by* $u\phi$ *where* ϕ *is a* S. P. *homeomorphism mapping* $\bar{D}$ *onto* $\bar{D}$.

Each of these operations on u yields an image C. E. with u. In cases (i) and (ii), Ru, and Δu are C. E. with u under the identity. Each of these operations has an inverse. The inverse of R is R. The operation Φ determined by the mapping ϕ has an inverse Φ^{-1} determined by ϕ^{-1}. If Δ is defined by $\Delta(z, t, u)$ as above, the inverse of Δ is defined by $\Delta(z, 1-t, u)$ for $u \varepsilon [U]$, $z \varepsilon \bar{D}$, and $0 \leqq t \leqq 1$. The operations R, Φ, Δ generate a group Ω.

The principal theorem can now be stated.

THEOREM 7.1. *Each element* $u \varepsilon [U]$ *has the form* ωU *where* ω *is an element in the group* Ω.

Given $u \varepsilon [U]$, U is C. E. with u under some mapping ϕ. Hence $u\phi$ is C. E. with U under the identity. In accordance with (α), $R_0 U$ is positively C. E. with $u\phi$, where $R_0 = R$, or the identity. We set $u\phi = \Phi u$ and introduce the deformation

$$\Delta(z, t, R_0 U) = (1-t)(R_0 U)(z) + t(\Phi u)(z) \qquad (0 \leqq t \leqq 1). \tag{7.1}$$

For $0 < t < 1$ the two terms in the right member define positively C. E. functions whose sum is in $[U]$ by virtue of (β). As a consequence of (7.1)

$$\Delta R_0 U = \Phi u \to \Phi^{-1} \Delta R_0 U = u \tag{7.2}$$

This establishes the theorem.

The first relation in (7.2) has the following meaning.

COROLLARY. *If* $u \varepsilon [D]$ *and* $U \varepsilon [D]$ *are* C. E. *under the identity there exists a continuous deformation of* u *through elements in* $[U]$, *each* C. E. *with* U *under the identity, into one of the two elements* $\pm U$.

Part II. Pseudoconjugates.

8. Definition. Let u be pseudoharmonic in D. A pseudoharmonic function v such that $u + iv$ is interior in D will be called pseudoconjugate to u in D. If u is continuous on $\bar{D}$ and pseudoharmonic in D, a function v continuous in $\bar{D}$ and pseudoconjugate to u in D will be termed pseudoconjugate to u *in* $\bar{D}$. When v is pseudoconjugate to u in D a point $z_0 \varepsilon D$ is an ordinary or multiple point of u of index n, if and only if it has the same character relative to v. We shall simplify the problem of constructing pseudoconjugates by noting the following:

(β). *If v is pseudoconjugate to $u \varepsilon [D]$, and if ϕ is any* S. P. *homeomorphism of $\bar{D}$ onto itself, then $v\phi$ is pseudoconjugate to $u\phi$.*

This follows at once from the definition of a pseudoconjugate.

As a consequence of (β) to construct a pseudoconjugate to $U \varepsilon [D]$ one can use any domain E such that $\bar{E}$ is the topological image $\phi(\bar{D})$ of $\bar{D}$ under a S. P. homeomorphism ϕ, and replace U by $u \varepsilon [E]$ where $U = u\phi$. If then v is constructed pseudoconjugate to u, $v\phi$ is pseudoconjugate to U.

(γ). *Let u and v be pseudoharmonic functions in $\bar{D}$ such that $u + iv = f$ maps $\bar{D}$ topologigcally into the complex w-sphere. Then f is sense-preserving and v therefore pseudoconjugate to u if there exists a continuous* 1-*parameter family of topological mappings f^t $(0 \leqq t \leqq 1)$ of $\bar{D}$ onto the complex w-sphere such that $f^0 = f$ and f^1 is interior.*

This follows from the primitive definition of a S. P. topological mapping. In applying (γ) one can take f^1 as analytic, or if convenient as the identity.

9. Three special constructions of pseudoconjugates. By virtue of the remarks of § 8 the three special constructions of pseudoconjugates now to be given have wide application. We refer to the complex plane of $z = x + iy$.

I. *Let E_1 be the square $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$. On E_1 let u be the pseudoharmonic function with the values $ax + b$, $a > 0$. Let $h(y)$, $0 \leq y \leq 1$, be continuous and strictly increasing. Then a function v with values*[7] *$v(z) = h(y) + x$ is pseudoconjugate to u.*

Since $u + iv$ clearly defines a topological mapping of the z-plane into another complex plane the only point of difficulty is in proving that $u + iv$ is sense preserving. In accordance with (γ) of § 8 this is established by deforming $h(y)$ through a continuous 1-parameter family of strictly increasing functions into the identity. When $v(z) = y + x$, $u + iv$ is clearly sense preserving and I follows.

---

[7] The x is added to make the boundary values of v strictly monotone when $y = 0$, or 1.

II. *Let E_2 be the semi-disc: $x \leqq 0$, $|z| \leqq 1$ with diameter d on which $x = 0$. Let u be the pseudoharmonic function with values $ax + b, a > 0$. Let $h(y)$ be continuous and strictly increasing. Then a function v with values $v(z) = h(y)$ is a pseudoconjugate of u.*

III. *On the disc E_3: $|z| \leqq 1$ the function v with values y is pseudoconjugate to the function u with values $ax + b$, where $a > 0$.*

The pseudoconjugates v defined in I, II and III lead to more general constructions as follows: Let $R \subset D$ be a Jordan domain such that $\bar{R}$ is mapped by a S. P. homeomorphism ϕ onto $\bar{E}_i$, i $= 1, 2, 3$. Then the functions $U = u\phi$ and $V = v\phi$ are defined over $\bar{R}$ and V is a pseudoconjugate of U. More definitely we limit R to subregions R_i of D as follows:

Case I. $\bar{R}_1$ shall be bounded in $\bar{D}$ by two disjoint arcs of βD and by two simple disjoint non-intersecting arcs p and q, intersecting βD in a finite number of points including the end points of p and q.

Case II. $\bar{R}_2$ shall be bounded in $\bar{D}$ by an arc of βD and by a simple arc d, in D except for a finite number of points, including d's end points.

Case III. R_3 shall be identical with D.

By virtue of the constructions I, II and III functions U pseudoharmonic on R_i for which pseudoconjugates V exist can be characterized together with V as follows (brackets indicate alternative):

Class I. Suppose that a $U \varepsilon [R_1]'$ (See § 4) assumes its minimum (maximum) and its maximum (minimum) respectively at each point of the arcs p, q of βR_1, is strictly monotone on the two complementary arcs of βR_1 and has no multiple points at any point of $\bar{R}_1$. A pseudoconjugate V of such a U always exists with values which are prescribed on p and strictly decreasing [8] (increasing), which has no multiple points in $\bar{R}_1$ and is strictly monotone on βR_1 except for a minimum (maximum) and maximum (minimum) at the final points of p and q respectively.

Class II. Let a $U \varepsilon [R_2]'$ assume its absolute maximum (minimum) at each point of the arc $d \varepsilon \beta R_2$, assume its absolute minimum (maximum) at a point $P \varepsilon \beta R_2$, possess no other extreme boundary points and no multiple points. A pseudoconjugate V of U then exists without multiple points in $\bar{R}_2$,

[8] Sense of increase of functions on βR_i are with respect to an independent variable moving on βR_i in a counter-clockwise sense. The arcs p and q are to be sensed counter-clockwise on βR_1 for this purpose.

with prescribed strictly increasing (decreasing) values in d and strictly decreasing (increasing) values in $\beta R_2 - d$.

Class III. Let $U \varepsilon [D]$ have no multiple points and $U \mid \beta D$ just two extreme points. Then there exists a pseudoconjugate V of U of the same character as U such that the extreme points of U and V appear in βD in the circular order min U, min V, max U, max V.

In establishing the existence of the pseudoconjugates V one begins by showing that the respective functions U in Classes I, II, III are strictly C. E. with $\pm u$ in I, II, III, making use of μ-lengths along the U-positive sensed level arcs (Cf. §§ 5, 6) to obtain the appropriate mapping ϕ. The existence of pseudoconjugates to functions U in Classes I, II, III is all that we shall need to establish the existence of a pseudoconjugate to an arbitrary $u \varepsilon [D]$.

10. Secteurs and inner boundaries. A pseudoconjugate of an arbitrary $U \varepsilon [D]$ will eventually be constructed out of the special constructions of § 9. To that end D must be broken up into special regions X coming under Cases I, II, and III, with $U \mid X$ in Classes I, II, III respectively. We shall need several definitions.

If g is a sensed arc the corresponding unsensed arc will be denoted by $\mid g \mid$. We say that $\mid g \mid$ *carries* g.

The net $\mid N(U) \mid$. Let z_0 be an arbitrary multiple point of U. Let h be a U-arc issuing from z_0 and continued in the ordinary sense until βD or another multiple point is reached. Let $N(U)$ be the union of all such sensed U-arcs. Let $\mid N(U) \mid$ be the union of the corresponding unsensed arcs.

Right or left continuations. An arc $h_2 \varepsilon N(U)$ is termed the *right* (*left*) continuation of an arc $h_1 \varepsilon N(U)$ if h_1 terminates in the initial point P of h_2 and if $\mid h_1 \mid$ and $\mid h_2 \mid$ are the right and left rays respectively (left and right rays) as seen from P, bounding a sector in $S(P, U)$. A simple sensed arc $h'(h'')$ composed of a sequence of arcs of $N(U)$ is termed the *maximal right* (*left*) *continuation* of each of its subarcs in $N(U)$ if the second of any two successive arcs in $N(U)$ and $h'(h'')$ is the right (left) continuation of the first, and if $h'(h'')$ is a proper subarc of no arc with this property. These maximal right (left) continuations are obviously simple U-arcs with end points on βD.

Inner boundaries. The preceding maximal right (left) continuations $h'(h'')$ if reversed in sense are maximal left (right) continuations. An unsensed arc k identical with $|h'|$ or $|h''|$ will separate D into two or more regions the closure of one and only one of which (termed a *secteur* K) will contain k and no elements of $|N(U)|$ incident with k other than elements of $|N(U)| \cap k$. Observe that $D - \bar{K}$ is not a secteur. The secteur K is bounded by k and a unique arc of βD. We term k an *inner boundary.* It is clear that K determines and is determined by its inner boundary k. Two inner boundaries either do not intersect or intersect in a point or arc of $|N(U)|$. If k' is a second inner boundary and if $k \cap k' \neq 0$, k' is in $D - K$. Of the regions into which k separates D, the secteur K determined by k is distinguished from the other components of $D - k$ by the following property: if λ and μ are two elementary arcs of $k \cap |N(U)|$ with a multiple point P as common end point, the sector of a canonical neighborhood of P bounded by λ and μ belongs to K. Every element $|h_1|$ in $|N(U)|$ belongs to just two inner boundaries, carrying the maximal right and left continuations of h_1.

Let z be a point not in $|N(U)|$. Of the connected regions into which $|N(U)|$ divides D let X be the region containing z. If βX includes an arc $|h_1| \varepsilon |N(U)|$, it is clear that it must include the carrier p of either the right or left continuation of h_1. The U-arc p is the inner boundary of some secteur K. Then either $X \subset K$ or else $X \cap K = 0$. The case $X \subset K$ prevails for the following reasons. If P is a multiple point of U in p then X intersects $S(P, U)$ near P in just one sector S' of $S(P, U)$ and p contains both arcs of $|N(U)|$ on the boundary of S' issuing from P. It is characteristic of $\bar{K}$ that, containing both of these arcs of $|N(U)|$, it contains the sector S' in a neighborhood of p. Thus $X \subset K$.

LEMMA 10.1. *The net* $|N(U)|$ *(assumed non-empty) divides* D *into a finite set of connected open regions* X *containing no multiple points of* U. *Each region* X *is either,*

(i) *the intersection* R *of two secteurs* K *and* G *possessing non-intersecting inner boundaries* p *and* q, *with* R *bounded by* p, q *and two disjoint arcs of* βD; *or*

(ii) *a secteur* K.

In Case (i), $U \mid \bar{R}$ *is in Class* I *of* § 9. *In Case* (ii), $U \mid \bar{K}$ *is in Class* II.

If X is bounded by p and an arc of βD, then X is a secteur K and Case (ii) alone occurs. In this case it is clear that $U \mid \bar{K}$ is in Class II of § 9.

If Case (ii) does not arise, βX contains a second inner boundary q, the inner boundary of some secteur G. As before $G \supset X$. Moreover $p \cap q = 0$;

otherwise q would include arcs of $|N(U)|$ in the complement of $\bar{K}$, which is impossible since $\bar{X} \subset \bar{K}$. We see that βX includes at most a finite set of disjoint maximal connected U-arcs. But $U \mid \bar{X}$ has no multiple points, so that it follows as in § 6 (α) and (β) that βX contains just two disjoint inner boundaries p and q, and $U \mid \bar{X}$ is in Class I of § 9.

11. The general construction of pseudoconjugates. If $U \varepsilon [D]$ has no multiple points at all, U is in Class III of § 9 (Cf. (β) of § 6), and a pseudoconjugate V of U is imemdiate. We assume therefore that U has at least one multiple point.

It will be convenient to term the region R arising in Case (i) of Lemma 10. 1 a *secteur band.* By means of the net $|N(U)|$, $\bar{D}$ has been decomposed into a finite number of secteur bands and secteurs on which U is in Class I and Class II respectively. To construct pseudoconjugates of U over the closure of any one of these regions X so as to yield a resultant continuous pseudoconjugate V of U one must progressively assign boundary values of V (termed V-values) along preferred U-arcs of βX as in the constructions of § 9.

V-values. If z_1 is an ordinary point of U on a U-arc g and if g is sensed U-positively [U-negatively] (see § 5) then a pseudoconjugate V of U *must* decrease [increase] along g near z_1 in order that $U + iV$ may be sense preserving.

To construct a function V pseudoconjugate to U we prefer an arbitrary one of the secteurs K or secteur bands R into which D is separated by $|N(U)|$.

Case a. In the case of K, continuous strictly monotone values (termed V-values) will be arbitrarily assigned along the inner boundary k of K so that these values increase or decrease in the unique sense possible for a pseudoconjugate of U. Recall that $U \mid \bar{K}$ is in Class II of § 9. (Lemma 10. 1.) Hence U possesses a pseudoconjugate V over $\bar{K}$ extending the V-values just assigned along k. The resulting $V \mid \bar{K}$ is strictly monotone over $\beta D \cap \beta K$.

Case b. If the preferred region is a secteur band R we choose one of the inner boundaries k in βR and assign V-values along k as in Case a. In accordance with Lemma 10. 1, $U \mid \bar{R}$ is in Class I of § 9. There accordingly exists a pseudoconjugate V of U over $\bar{R}$ extending the V-values assigned along k. Along the second inner boundary k' in βR, V, as constructed over $\bar{R}$, gives strictly monotone values, increasing or decreasing in the unique manner possible for a pseudoconjugate of $U \mid \bar{R}$.

Let X be any secteur or secteur band in the decomposition of D by $|N(U)|$ such that βX intersects βK, Case a, or βR, Case b, in an arc. This intersection must be along a single connected arc g, since no U-arc can be closed in $\bar{D}$. Let Σ denote K or R according as Case a or Case b arose in the first construction. Let $k \supset g$ be the inner boundary of X which intersects $\beta\Sigma$. Note that $g \neq k$. Recall that V-values have already been constructed on g. V-values will now be assigned continuously on the residual arc of k so that the V-values are strictly monotone over all of k. A function V pseudoharmonic over X can then be constructed as in Case a or Case b extending the values given on k.

In general let Σ be the closure of the union of the secteurs and secteur bands on which V has already been constructed. As in the second stage of the process let X represent any secteur or secteur band in the decomposition of D such that $\beta X \cap \beta\Sigma$ is an arc g. Let $k \supset g$ be the inner boundary of X which intersects $\beta\Sigma$. In the general case it is possible that $k = g$. In case $k = g$, V-values have already been constructed along k. In case $k \neq g$, V-values are assigned along k as in the second stage. A function V is then constructed over X as in the second stage. This process is continued until V is constructed over all of $\bar{D}$.

It may be remarked that the construction of V is such that $V \mid \beta D$ is strictly monotone over each of the arcs of βD into which βD is separated by $|N(U)|$.

It remains to prove that $f = U + iV$ is interior over D.

It is interior by construction at each point $z_0 \varepsilon D$ not in the net $|N(U)|$. Let z_0 then be in $|N(U)| \cap D$ but not a multiple point of U. Without loss of generality we can suppose that in a sufficiently small neighborhood of z_0, $U(z) = x + c$, for this would be true of a composite function $U\phi$ where ϕ was a suitable S. P. homeomorphism of a neighborhood of z_0. As explicitly constructed V is continuous and strictly decreasing along each U-arc (with U-positive sense) sufficiently near z_0, so that it is clear that f is sense preserving and topological in a sufficiently small neighborhood of z_0.

Suppose finally that $z_0 \varepsilon D$ is a multiple point of U. If $e > 0$ is sufficiently small and N_e is the neighborhood $(|z - z_0| < e)$ of z_0, $f \mid N_e$ satisfies the conditions of Lemma 2.1 and is accordingly interior at z_0. The following theorem is accordingly proved.

THEOREM 11. 1. *Corresponding to an arbitrary $U \varepsilon [D]$ there exists a pseudoconjugate $V \varepsilon [D]$ of U pseudoharmonic over D and continuous over $\bar{D}$.*

12. The continuation of interior transformations. To complete our results the following theorem is needed:

THEOREM 12. 1. *An interior mapping f of a Jordan domain D into the w-plane, continuous on $\bar{D}$, which send βD into a locally simple curve g* [M., p. 62] *can be extended to an interior transformation over a Jordan domain $\Delta \supset \bar{D}$.*

To establish this theorem a Riemann ribbon Σ spread over the w-plane and bearing a curve g' which is simple on Σ and projects into g will be constructed. The more precise statement is as follows:

LEMMA 12. 1. *Given a locally simple sensed curve g in the w-plane there exists an interior mapping Z of the form $w = Z(z)$ of a domain*

$$(12.1) \qquad a < |z| < a' \qquad (0 < a < 1) \qquad (aa' = 1)$$

of the z-plane into the w-plane which sends the unit circle ($|z| = 1$) in counter-clockwise sense into g, and which is such that for some constant $e > 0$ and for each subdomain of (12. 1) *for which $\theta_0 - e \leqq \arg z \leqq \theta_0 + e$ the mapping Z is topological.*

The Riemann surface of Z^{-1} is the Riemann ribbon to which reference was made.

To prove this lemma let $\cdots w_{-2}, w_{-1}, w_0, w_1, w_2, \cdots$ be a cyclic sequence of points on g with $w_i = w_{i+n}$ for some $n > 3$ and all i, so chosen that not only the subarcs $g_i = [w_i, w_{i+1}]$ but also the subarcs

$$(12.2) \qquad [w_i, w_{i+1}, w_{i+2}, w_{i+3}]$$

of g are simple. Let k_i $(i = 0, \pm 1, \pm 2, \cdots)$ be a simple Jordan arc in the w-plane such that k_i intersects the subarc $[w_{i-1}, w_i, w_{i+1}]$ of g in w_i alone, where w_i is an interior point of k_i and where $k_i = k_{i+n}$ for all i. Because of the simplicity of the arc (12. 2) we may assume that k_i intersects neither k_{i-1} nor k_{i+1}. There exists a sense-preserving topological mapping T_r of the z'-plane into the w-plane which sends the arcs $[y' = 0, 0 \leqq x' \leqq 1]$ into g_r and the arcs $[x' = 0, -1 \leqq y' \leqq 1]$ and $[x' = 1, -1 \leqq y' \leqq 1]$ into k_r and k_{r+1} respectively. The mapping T_r parametrizes k_r and k_{r+1} in terms of y'. We can suppose $T_0, T_{\pm 1}, T_{\pm 2}, \cdots$ successively chosen so that T_r and T_{r+1} give the same parameterization to k_{r+1}, and so that $T_i = T_{i+n}$ for all i. In particular T_{n-1} can be chosen so as to give the same parameterization to $k_0 = k_n$ as does T_0, since the w-plane is orientable. The condition of simplicity for the arcs (12. 2) has the following consequence. If $b > 0$ is sufficiently small, then for z' and z'' in the rectangle $H: 0 \leqq x' \leqq 1$, $-b \leqq y' \leqq b$, $T_r(z') = T_{r+1}(z'')$ if and only if $z' - z'' = 1$ and if z'' is pure imaginary or null. That is for z' and z'' in H, the mappings T_r and T_{r+1} coincide only in their mapping of $-b \leqq y' \leqq b$ onto k_{r+1}.

Let T now map the strip $[-\infty < x' < \infty, -b \leqq y' \leqq b]$ into the w-plane with T defined by setting

$$T(z' + r) = T_r(z') \qquad (r = 0, \pm 1, \pm 2, \cdots)(z' \varepsilon H). \tag{12.3}$$

Under T the map of the x'-axis is an unending locally simple sensed curve which reduces to g by virtue of the relation $T(x' + n) = T(x')$. If e_1 is a sufficiently small positive constant and if c is an arbitrary value of x' the mapping T taken over the rectangle

$$c - e_1 \leqq x' \leqq c + e_1, -b \leqq y' \leqq b, \tag{12.4}$$

is sense-preserving and topological. For z' in the domain of definition of T set $z = \exp(2\pi i z'/n)$ and subject to this relation set $Z(z) = T(z')$. The resulting mapping Z has the properties affirmed in the lemma.

Proof of Th. 12. 1. It follows from Lemma 2. 2 that f can be extended as an interior transformation over the neighborhood of each point z_0 of βD neighboring which in $\bar{D}$, f fails to be topological. Since there is at most a finite set of such points on βD [M., p. 85] one can suppose D replaced by a Jordan domain $\Delta \supset D$, over whose closure $\bar{\Delta}$, f can be extended so as to be interior in Δ, continuous in $\bar{\Delta}$ and topological in some neighborhood (relative to $\bar{\Delta}$) of each point of $\beta\Delta$. Without loss of generality we can finally suppose that f is interior over a unit disc D: $(|z| < 1)$, continuous on $\bar{D}$ and locally topological in some neighborhood (relative to $\bar{D}$) of each point of βD.

Such an f maps some neighborhood N (relative to $\bar{D}$) of each point z_0 of βD in a topological manner onto a subset of a neighborhood of the point $f(z_0)$ of g. The points in $f(N)$ are either in g or locally on one side of g. Turning to Lemma 12. 1 let Σ_a be the Riemann surface of Z^{-1} over the w-plane, understanding that $Z(z)$ is defined only over (12. 1). Under Z, βD goes into a curve g' on Σ_a where g' is simple on Σ_a and projects onto g. If the constant $b > 0$ is sufficiently small f maps the domain E_b: $[1 - b \leqq |z| \leqq 1]$ in a topological manner into the closure of one of the two sets, say G_a, into which Σ_a is separated by g'. We suppose b so restricted.

Understanding that $f(z)$ is a point in $\bar{G}_a$ and not merely a point in the w-plane it is clear that $Z^{-1}f(z) = \Phi(z)$ is uniquely defined for $z \varepsilon E_b$ and that Φ maps E_b topologically into the z-plane, thereby mapping the circle

($|z|=1$) onto itself topologically. We extend Φ over the reflection E'_b of E_b in the unit circle by setting $\Phi(z)=(\Phi(z'))'$ where the primes denote reflection in the unit circle. Noting that $Z\Phi(z)=f(z)$ for $z \varepsilon E_b$, we extend f over E'_b by requiring this relation to hold over E'_b. Since Z is locally topological over E'_b, f is likewise locally topological over E'_b and hence interior over $E_b \cup E'_b$. The proof of Theorem 12. 1 is complete.

13. Existence of a harmonic function contour equivalent to a given $u \varepsilon [D]$. The present section is devoted to the proof of the following result.

THEOREM 13. 1. *If D is a Jordan domain and u a pseudoharmonic function in $[D]$ there exists a harmonic function U in $[D]$ such that U is strictly* C. E. *to u.*

In order to apply the results of § 12 we introduce the following lemma.

LEMMA 13. 1. *If D is a Jordan domain and $u \varepsilon [D]$ there exists a Jordan domain $\Delta \supset D$ and an extension u^* of u with $u^* \varepsilon [\Delta]$ such that u^* has no multiple points in $\beta\Delta$.*

Let $a \varepsilon \beta D$ be a multiple point of u to which, as a limiting point, m u-arcs tend. Without essential loss of generality suppose that $u(a)=0$. In the w-plane consider the harmonic function H with values $H(w)=\mathfrak{R}\, w^m$. ($\mathfrak{R}$ denotes real part.) Let W be the neighborhood of $w=0$ in which $|w|<e$. Let W be divided into the semi-disc W': $|w|<e$, $0<\arg w<\pi$ and the complement of the latter relative to W. Let $\mathfrak{D}$ be the semi-disc defined by $|w|<r$, $0<\arg w<\pi$ $(r>e)$. The function $H \mid \bar{W}'$ has m level arcs tending to $w=0$ in W'. Moreover $H \mid \bar{W}'$ is strictly C. E. to $\pm u$ over a canonical neighborhood of $z=a$ in $\bar{D}$. This contour equivalence can be extended as a S. P. homeomorphism ψ of the w-plane onto the z-plane, mapping $w=0$ onto $z=a$, $\mathfrak{D}$ onto D and, $\beta\mathfrak{D}$ onto βD and such that

$$u(\psi(w))=\sigma H(w) \qquad (w \varepsilon \bar{W}', \sigma=\pm 1),$$

provided e is sufficiently small.

We extend u as a pseudoharmonic function across βD near $z=a$ by setting $u(\psi(w))=\sigma H(w)$ $(w \varepsilon W)$. Now consider a Jordan domain D' obtained from D by a modification of βD on a short arc containing $z=a$ in its interior. This modification corresponds under ψ to a change conveniently made in the w-plane, recalling that the segment g of the real w-axis from $w=-e$ to $w=e$ is sent by ψ into an arc of βD containing $z=a$. Let the arc of g from $w=-e/2$ to $w=e/2$ be replaced by a circular arc γ on

$|w| = e/2$, lying in the lower half w-plane apart from its end points. If g' is the open arc thereby replacing g, $H \mid g'$ has $m - 1$ extreme points, on γ, not multiple points of $H \mid W$. Through every other point of g' just one level arc of H enters the domain bounded by g' and the arc $|w| = e$, $0 \leq \arg w \leq \pi$. On the arc corresponding to g' under ψ in the z-plane there is thus no multiple point of u. Making corresponding extensions of u and modifications of the boundary in a neighborhood of every other multiple point of u in βD, we obtain the desired domain Δ. The extension u^* of u to Δ is clearly in $[\Delta]$ and has no multiple point in $\beta\Delta$.

Now let us suppose that u in Th. 13. 1 has been extended to the Jordan domain Δ so as to have these properties. Let v be a pseudoconjugate of u in $\bar{\Delta}$ constructed according to the prescriptions of §§ 9, 11. Then $u \mid \beta\Delta$ and $v \mid \beta\Delta$ have no common extreme point. Indeed, if $|N(u)|$ now denotes the net of u as extended to Δ, the extrema of $v \mid \beta\Delta$ occur at most at points of $|N(u)|$ in $\beta\Delta$, while the extrema of $u \mid \beta\Delta$ occur at most on the open arcs into which $\beta\Delta$ is divided by the points of $|N(u)| \cap \beta\Delta$. Thus $f = u + iv$ is interior at the points of Δ and maps $\beta\Delta$ on a locally simple curve.

By Th. 12. 1 f can be extended to a domain Σ containing $\bar{\Delta}$ (and so $\bar{D}$) in its interior. The Riemann surface which is the image of Σ under f can be mapped conformally, say by a function F, into the finite z-plane. Corresponding to D under Ff we obtain a Jordan region E. Next E can be mapped conformally, say by a function G, onto D, and by a well-known result the mapping can be extended to a homeomorphism between $\bar{E}$ and $\bar{D}$. The compound mapping GFf is thus a S. P. homeomorphism ϕ of D onto itself such that $f\phi^{-1}$ is a regular function on D. Thus the function $U = u\phi^{-1}$ is harmonic on D, continuous on $\bar{D}$ and clearly strictly C. E. to u.

It should be remarked that the proof in the last paragraph could also be carried out by first extending f to a domain containing Δ as in the first paragraph of the proof of Th. 12. 1 so as to eliminate partial branch elements [M., pp. 83, 85], and then using the above conformal mapping theorem for closed Jordan domains in a neighborhood of each boundary point relative to the closed domain. In this way one would not require the full force of Th. 12. 1.

14. Model functions U and nets $|N(U)|$. Suppose that D is the domain $(|z| < 1)$. Recall that the net $|N(U)|$ of U is defined in § 10. We seek a topological model for the net of U under S. P. homeomorphisms $T(D)$ of $\bar{D}$ onto $\bar{D}$. Recall that the domain D may be regarded as a hyperbolic plane with circles orthogonal to $(|z| = 1)$ as its straight lines. The closure in $\bar{D}$ of any such arc will be called an H-line.

THEOREM 14. 1. *There exists a topological model in $\bar{D}$ for the net $|N(U)|$ of $U \varepsilon [D]$ which is the union of a finite set of H-lines in $\bar{D}$.*

Let g be the continuation in D in the sense of § 2 of any one of the elements $h \varepsilon N(U)$. We term $|\bar{g}|$ a *complete arc* in $|N(U)|$. Two points of $\bar{D}$ will be said to be *cofinite* if both are in D or both in βD.

Set $N_0 = |N(U)|$. By a *vertex* of N_0 is meant a multiple point of U or a point of N_0 in βD. By an *element* of N_0 is meant any arc of N_0 whose end points are vertices of N_0 but which carries no other vertices of N_0. Given two vertices P_r, P_s of N_0 let $N_0(P_r, P_s) = \infty$ if P_r, P_s are not connected on N_0. If P_r and P_s are connected on N_0 let $N_0(P_r, P_s)$ be the minimum number of elements of N_0 which is necessary to traverse to pass from P_r to P_s. It is clear that $N_0(P_r, P_s) = N_0(P_s, P_r)$. We term $N_0(P_r, P_s)$ the *N_0-distance* from P_r to P_s.

We suppose that the vertices of N_0

$$P_1, P_2, \cdots, P_n \tag{14. 0}$$

have been ordered as follows. Choose P_1 arbitrarily among vertices of N_0. Choose next the vertices $P_i(i > 1)$ for which $N_0(P_1, P_i)$ is finite taking these vertices in the order of magnitude of $N_0(P_1, P_i)$, and arbitrarily when the N_0-distances from P_1 are equal. Suppose that $P_1, P_2. \cdots, P_r$ have been so ordered and that this set includes every vertex connected on N_0 to any member of the set. Choose P_{r+1} arbitrarily among the remaining vertices of N_0 (if there are any). Follow P_{r+1} by the vertices $P_j(j > r+1)$ which are connected to P_{r+1} on N_0 taking these vertices in the order of magnitude of the numbers $N_0(P_{r+1}, P_j)$ or arbitrarily if the N_0-distances from P_{r+1} are equal. This process will suffice to order the vertices of N_0.

There exists a S. P. homeomorphism T_1 of $\bar{D}$ onto $\bar{D}$ with $T_1(P_1) = P_1$ under which each complete arc in N_0 meeting P_1 has an image which is an H-line. This is readily established on using the Jordan separation and Schoenflies mapping theorems.

Proceeding inductively suppose that for $0 < r < n$ there exists a S. P. homeomorphism T of $\bar{D}$ onto $\bar{D}$ such that the image net $T(N_0) = N$ and the image vertices $T(P_i) = Q_i$, $i = 1, \cdots, n$, have the property that each complete arc h of N meeting the set

$$Q_1, \cdots, Q_r \tag{14. 1}$$

is an H-line. This has been established for $r=1$. Two cases are to be distinguished.

Case I. Q_{r+1} is connected to Q_r on N.

Case II. Not Case I.

Case I. In this case some point Q_j, $j \leqq r$, and Q_{r+1} determine an H-line b of N. There exists a point Q on b cofinite with Q_{r+1}, with Q_j, Q_{r+1}, Q in the order written on b, and with Q so near βD that there are H-lines meeting Q which do not meet the H-lines h of N except b. (If Q_{r+1} is in βD we take $Q = Q_{r+1}$.) There will then exist a homeomorphism T^1 of $\bar{D}$ onto $\bar{D}$ for which $T^1(Q_1) = Q_1, \cdots, T^1(Q_r) = Q_r$ and under which the H-lines of N meeting the set (14.1) are arc-wise invariant, while the image under T^1 of each complete arc of N meeting Q_{r+1} is an H-line meeting Q. The above inductive hypothesis made on the sets (14.0), (14.1) and T, is now seen to be satisfied by the sets $P_1, \cdots, P_n$, $Q_1, \cdots, Q_r$, Q, and T^1T; in particular $T^1T(P_{r+1}) = T^1(Q_{r+1}) = Q$.

Case II. Let N^1 be the subset of complete arcs of N meeting the set (14.1). The point Q_{r+1} lies in a secteur or secteur band R of N^1. One replaces each complete arc of N meeting Q_{r+1} by an H-line through Q_{r+1}. Then the inductive hypothesis holds for suitable T and r replaced by $r+1$.

The theorem follows by induction.

We turn to an inverse problem: What are the characteristics of a set of H-lines in $\bar{D}$ that it may serve as the net of some $U \varepsilon [D]$? The answer is in terms of the following definition.

Definition. *We admit any finite set N^* of H-lines in D each of which intersects at least one other H-line and which does not include the entire boundary of any domain in D.*

According to Theorem 14.1 there exists a topological model of the net $|N(U)|$ of a $U \varepsilon [D]$ which is a set N^* of H-lines as admitted above. Conversely we have the theorem

Theorem 14.2. *Corresponding to any admissible set N^* of H-lines there exists a $U \varepsilon [D]$ for which $N^* = |N(U)|$.*

It is clear that N^* separates $\bar{D}$ into a finite number of secteurs or secteur bands R. The proof of the existence of a $U \varepsilon [D]$ for which $N^* = |N(U)|$ is similar to the proof of the existence of a pseudoconjugate $V \varepsilon [D]$ as given in § 11. One adds the regions R successively so as to always keep a simply connected domain. After the first step, U is given by the previous construction of U as constant on a single arc of $\beta R \cap N^*$. In the case of a secteur R, U is first extended as a constant c over all of $\beta R \cap N^*$ and then extended over R with $U < c$ or $> c$ according as $U > c$ or $< c$ on the secteurs or secteur bands with which βR has arcs in common. No inconsistency can appear as a consequence of this demand at a vertex of N^* in βD nor at any interior vertex P, since there is always an even number of arcs of N^* incident with P. In the case of the adjoining of a secteur band R the value c on one boundary arc of R in N^* is determined by the previous construction of U. The value c' on the other boundary arc of R in N^* is arbitrary subject to one of the conditions $c' < c$, or $c' > c$ uniquely determined by the previous construction.

The theorem follows.

The preceding suggests a weaker form of Th. 4. 1.

THEOREM 14. 3. *A necessary and sufficient condition that u_1 and u_2 in $[D]$ with nets N_1 and N_2 respectively be contour equivalent under some* S. P. *homeomorphism ϕ of $\bar{D}$ onto $\bar{D}$ is that there exist a* S. P. *homeomorphism t of $\bar{D}$ onto $\bar{D}$ under which $t(N_1) = N_2$. If ϕ exists one can take $t = \phi$. If t exists one can take ϕ as an extension of $t \mid N_1$.*

If ϕ exists then $t = \phi$ satisfies the condition of the theorem. If t exists it is clear that in Th. 4. 1, $M(u_1)$ is similar to $M(u_2)$ under $t \mid N_1$ and in accordance with Th. 4. 1 one can take ϕ as an extension of $t \mid N_1$.

THE JOHNS HOPKINS UNIVERSITY AND
THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] M. Morse and M. Heins, "Deformation classes of meromorphic functions and their extensions to interior transformations," *Acta Mathematica*, vol. 79 (1947), pp. 51-103.

[2] James A. Jenkins, "On the topological theory of functions," to appear in the *Canadian Journal of Mathematics*.

THE ORTHOGONAL GROUP IN HILBERT SPACE.*

By Calvin R. Putnam and Aurel Wintner.

1. Let $\mathfrak{R}$ denote the real Hilbert space (in its realization in terms of vectors $x = \{x_k\}$ the components of which are real numbers x_k satisfying $|x| \equiv (\Sigma x_k^2)^{\frac{1}{2}} < \infty$). If A is a real, bounded matrix or, equivalently, a linear (distributive and continuous) operator transforming every point, x, of $\mathfrak{R}$ into a point, Ax, of $\mathfrak{R}$, let $|A|$ denote the least upper bound of the vector length $|Ax|$ when x varies over the unit sphere, $|x| = 1$, of $\mathfrak{R}$.

If an infinite matrix is real and unitary (hence bounded), it will be called an orthogonal matrix, O. Let Ω denote the metric space in which the points are the orthogonal matrices, O, and on which the distance is defined to be $|O_1 - O_2|$ (with the above meaning of $|A|$ for the difference, A, of two orthogonal matrices). Since $|O| = 1$ holds for every O, no distance $|O_1 - O_2|$ in Ω can exceed 2.

In particular, if I denotes the unit matrix, then

$$|O - I| \leqq 2 \tag{1}$$

holds for every orthogonal matrix. Let a point, O, of Ω be said to lie on the "boundary," $[\Omega]$, or in the "interior," $\Omega - [\Omega]$, of Ω according as the sign of equality does or does not take place in (1); so that

$$[\Omega]: \quad |O - I| = 2; \quad \text{i.e.,} \quad \Omega - [\Omega]: \quad |O - I| < 2. \tag{2}$$

2. Let Ω_0 denote the set of those matrices which are representable in the form $e^S (= I + S + \frac{1}{2}S^2 + \cdots)$, where S is some bounded, real, skew-symmetric matrix. It is easily verified that, for every such S, the matrix e^S is orthogonal, but the converse is not true (cf. (I) below); so that Ω_0 is a proper subset of Ω. A point, O, of Ω will be called a rotation or a reflection according as it is in Ω_0 or in its complement, $\Omega - \Omega_0$. This nomenclature is suggested by the circumstance that a *finite* real, orthogonal matrix is well-known to be of positive or of negative determinant $(= \pm 1)$ according as it is or is not representable as the exponential of a real, skew-symmetric matrix.

While no determinants are available for the matrices of Ω, it is clear from the above definition of Ω_0 that, if O is any orthogonal matrix,

$$ORO^{-1} \text{ is a rotation if } R \text{ is,} \tag{3}$$

* Received April 13, 1951.

and that

(4) R^{-1} is a rotation if R is.

On the other hand, it will be seen in § 11 and § 13 that, in contrast to what holds for finite matrices,

(5) R_1R_2 can be a reflection if R_1, R_2 are rotations,

and that

(6) R_1R_2 can be a reflection if R_1, R_2 are reflections.

It should be noted that (6) is not implied by (5), nor (5) by (6). On the other hand, since the orthogonal matrices form a group, (3) is equivalent to the statement that

(7) ORO^{-1} is a reflection if R is,

and (4) to the statement that

(8) R^{-1} is a reflection if R is.

It follows that either all or none of the $k!$ products of k orthogonal matrices $R_1, \cdots, R_k$ are rotations. Clearly, it is sufficient to prove this for $k = 2$, i. e., to show that if R_1R_2 is a rotation, then R_2R_1 is. But this follows by applying (3) to $R = R_1R_2$, $O = R_2$.

The following fact may also be mentioned:

(*) Neither Ω_0 nor $\Omega - \Omega_0$ is an open set (hence neither of them is a closed set) on Ω.

It is understood that these and all the subsequent topological notions refer to the topology determined by the $| O_1 - O_2 |$-metric on Ω.

3. In the case of finite matrices, the set of all orthogonal matrices breaks into two closed manifolds (those of determinant $+1$ and -1, respectively), which have no point in common. In contrast, it will follow from (I) and (II) below that Ω_0 and $\Omega - \Omega_0$ contain points which are in the closures of the respective complementary sets, $\Omega - \Omega_0$ and Ω_0. Actually (I) and (II) together will supply a characterization of all these points O of "confluence" (of either kind).

4. The preceding assertions concerning the "confluence' of Ω_0 and $\Omega - \Omega_0$ neither contain nor are contained in the following fact, proved in [2]:

(9) Ω is arcwise connected.

In the case of finite orthogonal matrices, the rotations are known to form a connected manifold, and the same is true of the reflections. For the case of infinite matrices, it was shown in [2] that

(10) Ω_0 is arcwise connected,

and it will be proved in § 29 that

(11) $\Omega - \Omega_0$ is arcwise connected.

If t is a real number, then tS is a real, bounded, skew-symmetric matrix whenever S is. Hence (10) is obvious; cf. [2]. In fact, if R is a rotation, say $R = e^S$, then $R(t) = e^{tS}$, where $0 \leqq t \leqq 1$, represents a continuous path contained in Ω_0. But this path begins $(t = 0)$ at I and ends $(t = 1)$ at the given R.

5. Results corresponding to (9)-(11) will be proved for the "boundary" and for the "interior" of Ω, as defined by (2):

(12) $[\Omega]$ is arcwise connected

and

(13) $\Omega - [\Omega]$ is arcwise connected.

It will also be shown that, from the point of view of the arcwise connectivity of the boundary, $[\Omega]$, with the "center," I, of Ω, there are on $[\Omega]$ two types of points:

(A) There exist on $[\Omega]$ points O corresponding to which it is possible to find continuous paths $Q(t)$, $0 \leqq t \leqq 1$, connecting $O = Q(1)$ to $I = Q(0)$ in such a way that the given O is the only point of the path which is not in $\Omega - [\Omega]$.

(B) There exist on $[\Omega]$ points O which cannot be connected to I in the way specified under (A).

It should be noted that (A) is not implied by (13).

6. It will be easy to show that the "interior" of Ω contains none of the reflections and not all of the rotations, i. e., that

(14) $[\Omega] \supset \Omega - \Omega_0 \neq [\Omega]$.

The set of the rotations on the "boundary," i. e., the intersection, $[\Omega]\Omega_0$, of $[\Omega]$ and Ω_0, has quite an involved structure on $[\Omega]$. In fact, it turns out that, while

(15) $[\Omega]\Omega_0$ is arcwise connected,

it is not dense on $[\Omega]$ (or on Ω_0), and it is neither open nor closed on $[\Omega]$.

In case of a finite dimension number, all rotations form an invariant subgroup (of index 2) of all orthogonal transformations, and so any fixed co-set supplies a topological mapping of the space of all rotations on that of all reflections. In view of (5) or (6), this argument cannot be applied in the present case. But it will remain undecided whether Ω_0, nevertheless, is topologically equivalent to $\Omega - \Omega_0$. If it should be, then the "dual" results are mere corollaries; for instance, (11) then is equivalent to (10) and, therefore, trivial. It would, of course, be sufficient (but not necessary) to assure the existence of *some* reflection, say R_0, having the property that $R_0 O$ is a reflection or a rotation according as O is a rotation or a reflection.

The Spectral Characterization of Rotations and Reflections.

7. The proofs will depend on the spectral resolution of unitary matrices [3], pp. 268-277, applied in [1] to the real subgroup of the unitary group.

The purpose of this chapter is a spectral characterization of the rotations (hence, of the reflections as well). In view of (3) (or of (7)), it will always be allowed to assume that the orthogonal matrix to be considered is given in any of its normal forms which can be attained by orthogonal transformations.

In particular, it can be assumed that the contribution of the continuous spectrum (if any) has been split off by an orthogonal transformation. After such a transformation, every orthogonal matrix appears in exactly one of the forms.

$$(16) \qquad O = C; \qquad O = P_\infty; \qquad O = P_n \dot{+} C, \qquad 1 \leqq n \leqq \infty,$$

where C and P_∞ denote infinite orthogonal matrices having no point spectra and no continuous spectra, respectively, while if $n < \infty$, then P_n denotes a finite, n-rowed orthogonal matrix. It is understood that, whether $n < \infty$ or $n = \infty$ in the third of the cases (16), the symbol $\dot{+}$ means this: $x = 0$ is the only common point of those two linear subspaces of the x-space $\mathfrak{R}$ on which C and P_n operate.

If ϕ is any (real) angle, let $B = B(\phi)$ denote the binary matrix representing rotation by ϕ in a plane. Thus $B(0)$ is the two-rowed unit matrix, and $B(\pi)$ the negative of it. Hence, if the multiplicity with which -1 occurs in the spectrum of P_n is either finite and even (possibly 0) or infinite, then P_n is orthogonally equivalent to a matrix which, when denoted simply by P_n, is of the form

$$(17) \quad P_n = Q_n \text{ or } P_n = 1 \dot{+} Q_n, \text{ where } Q_n = B(\phi_1) \dot{+} B(\phi_2) \dot{+} \cdots, \qquad (n \leqq \infty),$$

and (17) must be replaced by

(18) $P_n = R_n$ or $P_n = 1 \dot{+} R_n$, where $R_n = -1 \dot{+} B(\phi_1) \dot{+} B(\phi_2) + \cdots$, $(n \leqq \infty)$,

if -1 occurs in the spectrum of P_n with a multiplicity which is finite and odd (the angles ϕ_m need not be distinct). It is understood that 1 and -1 in (17)-(18) represent the one-rowed unit matrix and its negative; that, whether $n < \infty$ or $n = \infty$, the number of the B-terms is $\frac{1}{2}n$ or $\frac{1}{2}(n-1)$ in (17) according as n is even or odd, and is $\frac{1}{2}(n-1)$ or $\frac{1}{2}(n-2)$ in (18) according as n is odd or even; finally, that the alternative cases are needed in (17) as well as in (18) in order to take care of matrices P_n, where $n \leqq \infty$, in which the multiplicity with which $+1$ occurs in the spectrum is not or is finite and odd.

8. It follows from (IV) in § 28 below that, if R is an orthogonal matrix not containing -1 in its point spectrum, and if $E(\lambda)$, where $0 \leqq \lambda \leqq 2\pi$, denotes the spectral matrix of $-R$, then the matrix

$$(19) \qquad R_t = \int_0^{2\pi} e^{it(\lambda-\pi)} dE(\lambda), \text{ where } 0 \leqq t \leqq 1,$$

is orthogonal and satisfies the functional equation $R_u R_v = R_{u+v}$. Since (19) also implies that $R_1 = -Re^{-i\pi} = R$, it follows that $R_{\frac{1}{2}} R_{\frac{1}{2}} = R$. Accordingly, if -1 does not occur in the point spectrum of an orthogonal matrix, R, then R is the square of some orthogonal matrix, $R_{\frac{1}{2}}$.

9. This fact will be combined with the following

LEMMA. *A matrix is a rotation if and only if it is the square of some orthogonal matrix.*

Since e^S is the square of $e^{\frac{1}{2}S}$, only the first of the two assertions of this Lemma needs a proof. But M. H. Martin has proved ([1], p. 590) that every orthogonal matrix, O, can be factored (not in a unique way) as follows:

$$(20) \qquad O = Te^S, \text{ where } TS = ST \text{ and } T^2 = I.$$

Here I is the unit matrix, and T, S are two real, bounded matrices the first of which is symmetric while the second is skew-symmetric. Since (20) implies that $O^2 = e^{2S}$, and since $2S$ is skew-symmetric, it follows that every O^2 is a rotation.

10. It will now be easy to prove the main theorem on rotations:

(I) *An orthogonal matrix is a rotation if and only if the multiplicity with which* -1 *occurs in its point spectrum is either finite and even* (*possibly* 0) *or* ∞.

First, let O be a rotation and suppose, if possible, that -1 occurs in its point spectrum with a finite, odd multiplicity. After an orthogonal transformation, it can be assumed that O is given in the form

$$O = \begin{pmatrix} -I^{2h+1} & 0 \\ 0 & R \end{pmatrix},$$

where I^{2h+1} is the $(2h+1)$-rowed unit matrix and R does not contain -1 in its point spectrum. Since O is a rotation, there exists a real, bounded skew-symmetric matrix S such that

$$O = e^S, \qquad S = \begin{pmatrix} A & B \\ -B' & C \end{pmatrix}$$

where the prime denotes the operation of transposition and A is a $(2h+1)$-rowed square matrix. It follows from $Se^S = e^S S$ that $RB' = -B'$. Hence $B' = 0$, and so

$$e^S = \begin{pmatrix} e^A & 0 \\ 0 & e^C \end{pmatrix}.$$

The last three formula lines imply that $e^A = -I^{2h+1}$. But this contains a contradiction, since $\det e^A > 0$ but $\det -I^{2h+1} < 0$. This proves the second assertion of (I).

In view of (16) and of the above Lemma, the proof of (I) will be complete if it is shown that every matrix C is the square of some orthogonal matrix, and that the same is true of every matrix (17). But every binary rotation, $B(\phi)$, is the square of such a rotation, $B(\frac{1}{2}\phi)$, while the one-rowed rotation matrix 1 is its own square. Consequently, it is sufficient to show that every C is the square of an orthogonal matrix. But C denotes an orthogonal matrix having a continuous spectrum only, hence -1 is surely not in the point spectrum of C, i. e., $R = C$ is of the type considered in § 8 before the Lemma and is therefore the square of the corresponding $R_{\frac{1}{2}}$.

11. It was shown in [2] that every reflection is the product of three rotations, and the question was raised whether two rotations would not always suffice. It will now be shown that (I) implies that the answer to this question is affirmative, and can even be refined as follows:

(†) *Every reflection can be represented as the product of two rotations; in addition, the latter can be chosen so as to be commutable.*

The assertion of this corollary, (†), of (I) is invariant under an arbitrary orthogonal transformation of a given reflection. It can therefore be assumed that if the latter is denoted by O and if k, where $0 \leqq k \leqq \infty$, denotes the multiplicity with which 1 occurs in the point spectrum of O, then $O = I^k \dot{+} R$, where I^k denotes the k-rowed unit matrix, 1 is not in the point spectrum of R, and either of the terms I^k, R may be absent. In the proof of (†), two cases have to be distinguished, according as $k < \infty$ or $k = \infty$.

If $k < \infty$, then R is infinite, and so the assertion of (†) follows by writing $O = I^k \dot{+} R$ as the product of the two matrices $I^k \dot{+} (-R)$, $I^k \dot{+} (-I^\infty)$. In fact, (I) shows that both of these matrices are rotations, since, by assumption, -1 occurs with the multiplicity 0 in the point spectrum of $I^k \dot{+} (-R)$, and with the multiplicity ∞ in the point spectrum of $I^k \dot{+} (-I^\infty)$. In the remaining case, where $k = \infty$, it is sufficient to write O as the product of $-O$ and $-I^\infty$, since -1 occurs in the point spectra of the latter two matrices with the multiplicity ∞. This proves (†).

Clearly, (†) and (10) imply (9). It is also clear that (5) follows from (†), and therefore from (I).

In contrast, (6) will depend on (II) below. It will remain undecided whether, corresponding to the refinement (†) of (5), it is possible to refine (6) to the statement that every reflection is the product of two reflections.

The Closures of the Rotations and of the Reflections.

12. For the sake of brevity, let an orthogonal matrix, O, be said to have a *pure vicinity* (*in* Ω) if there exists a positive $\beta = \beta(O)$ having the following property: Either every orthogonal matrix, Q, satisfying $|Q - O| < \beta$ is a rotation or every such Q is a reflection (according as O itself is a rotation or a reflection, $|Q - O| < \beta$ being satisfied by $Q = O$). Every orthogonal matrix is of one of two possible types, in accordance with the following theorem:

(II) *An orthogonal matrix, O, has a pure vicinity (in Ω) or is a cluster point both of rotations and reflections according as -1 is not or is in the essential spectrum of O.*

Corollary. *An orthogonal matrix, O, has a pure vicinity (in Ω) if and only if -1 is not in the essential spectrum of O.*

Here and in the sequel, the *essential spectrum* of an orthogonal matrix,[1] O, is meant to be the set of those λ-values which are either cluster points of the spectrum of O or are in the point spectrum of O with an infinite multiplicity (or both). For instance, $\lambda = 1$ is in the essential spectrum of I. Since the continuous spectrum of every O is a perfect set (unless it is vacuous), a λ-value can belong to the essential spectrum for any one, for any two, or for all three, of the following reasons: λ is in the continuous spectrum; λ is a cluster point of the point spectrum; λ is an eigenvalue of infinite multiplicity. Since O is an infinite matrix, its essential spectrum always contains at least one point. Needless to say, the essential spectrum is a subset of the spectrum (in fact, the latter is a closed set). The notions of the point spectrum, the continuous spectrum, etc. are meant, of course, in terms of the spectral theory of unitary matrices ([3], pp. 268-277).

13. Theorem (II), the proof of which will be lengthy, supplies a short proof of (6). To this end, let D denote a diagonal matrix which differs from the infinite unit matrix I only in that a single diagonal element of I is changed to -1. Thus $-D$ is a rotation containing -1 in its essential spectrum. It follows therefore from (II) that there exists a sequence of reflections $L_1, L_2, \cdots$ such that $L_m \to -D$ as $m \to \infty$ (the convergence refers to the metric of Ω). Similarly, since the negative, $-I$, of the unit matrix is, by (I), a rotation, and since -1 is in its essential spectrum, it follows from (II) that there exists a sequence of reflections $K_1, K_2, \cdots$ such that $K_m \to -I$. Consequently L_mK_m tends to $-D$ times $-I$, i. e., $L_mK_m \to D$. But (I) and the definition of D show that D is a reflection not containing -1 in its essential spectrum. It follows therefore from (II) that $L_mK_m \to D$ is possible only if L_mK_m is a reflection from a certain m onward, say for every $m \geqq j$. Hence, a pair of orthogonal matrices satisfying (6) follows by choosing $R_1 = L_j$, $R_2 = K_j$.

As another application of (I) and (II), the assertions of (*) at the end of § 2 will now be proved. First, it is clear from (I) that there exist both rotations and reflections for which -1 is in the essential spectrum. It follows therefore from (II) that neither the set of all rotations nor that of all reflections is an open or a closed set.

---

[1] In the Hermitian case, the idea and certain fundamental applications of essential spectrum are due to H. Weyl, *Rend. Palermo*, vol. 27 (1909), pp. 373-392, and *Mathematische Annalen*, vol. 68 (1910), p. 251.

14. There will now be collected the tools needed from spectral theory.

If H is any bounded, Hermitian matrix, then e^{iH} is unitary. Conversely, it was shown in [3], pp. 268-277, that every unitary matrix is representable in the form e^{iH}, even if the spectrum of H is restricted to the interval $0 \leqq \lambda \leqq 2\pi$, and even if $\lambda = 2\pi$ is restricted to be not in the point spectrum of H; furthermore, there belongs to every unitary matrix essentially one H subject to these restrictions. Thus if $E(\lambda)$ denotes the spectral matrix of H, then

$$\text{(21a)} \qquad E(\lambda - 0) = E(\lambda), \; -\infty < \lambda < \infty,$$

and

$$\text{(21b)} \qquad E(\lambda) = \| 0 \| \text{ if } \lambda \leqq 0, \text{ and } E(\lambda) = I \text{ if } \lambda \geqq 2\pi.$$

In his thesis, M. H. Martin has specified those $E(\lambda)$'s which satisfy (21a)-(21b) and have the property that, if H is the Hermitian matrix having the spectral matrix $E(\lambda)$, then the unitary matrix e^{iH} is real, i. e., orthogonal. The result of his reality discussion ([1], pp. 598-600) is as follows: e^{iH} is orthogonal if and only if

$$\text{(22a)} \qquad \bar{E}(+0) = E(+0), \qquad (E(0) = \| 0 \|),$$

and there exists a real constant a having the property that

$$\text{(22b)} \qquad \bar{E}(\lambda) + E(2\pi - \lambda) = aI \text{ for } 0 < \lambda < 2\pi$$

holds at those interior points λ at which $\bar{E}(\lambda) + E(2\pi - \lambda)$ is continuous. The bars in (22a)-(22b) denote complex conjugation, i. e., ordinary transposition, of the Hermitian matrix E.

15. This machinery will not be needed in the proof of the *first* assertion of (II), which proceeds as follows:

Suppose that O is an orthogonal matrix not containing -1 in its essential spectrum, and let k, where $0 \leqq k < \infty$, be the multiplicity with which -1 is in the point spectrum ($k = \infty$ is impossible, since -1 is not in the essential spectrum). It will be shown that if Q is any orthogonal matrix for which $| Q - O |$ is smaller than a certain positive number $\beta = \beta(O)$, and if $l = l(Q)$ denotes the multiplicity with which -1 occurs in the point spectrum of Q, then $l < \infty$ and $l \equiv k \pmod 2$. In view of (I), this will prove the first assertion of (II) (and even more, since l is claimed to be finite).

Let $x_1, \cdots, x_k$ denote a set of linearly independent eigenvectors of O belonging to -1, so that $Ox_h = -x_h$ (this x_h-set is vacuous if $k = 0$).

Since -1 is not in the essential spectrum of O, it is clear that $|(O+I)x|$ has a positive lower bound when x varies over those unit vectors which are orthogonal to the space spanned by these eigenvectors. Let such a positive lower bound be denoted by 2β. It will be shown that the $\beta=\beta(O)$ thus defined has the desired property.

In order to see this, let Q be any orthogonal matrix satisfying $|Q-O|<\beta$. Then -1 is not in the essential spectrum of Q and *at most* k eigenvalues[2] of Q are in the λ-circle $|\lambda+1|\leqq\beta$. For suppose the contrary. Then there exists a unit vector, x, which satisfies the inequality $|(Q+I)x|\leqq\beta$ and is orthogonal to the space spanned by the k eigenvectors, $x_1,\cdots,x_k$, of O considered above. But $|(O-Q)x|<\beta$, since $|x|=1$ and $|O-Q|<\beta$. It follows therefore from $|(Q+I)x|\leqq\beta$ that $|(O+I)x|<2\beta$. This contradicts, however, the definition of 2β as a lower bound.

If the rôles of O, Q are interchanged in the preceding proof, what follows is that Q has *at least* k eigenvalues in the λ-circle $|\lambda+1|\leqq\beta$. Consequently, Q has *exactly* k eigenvalues (and no essential spectrum) in this circle. But if λ is a complex eigenvalue of Q, then, since Q is orthogonal (hence real), the complex conjugate of λ is an eigenvalue of the same multiplicity as λ. Since the circle $|\lambda+1|\leqq\beta$ is bisected by the real axis and since the spectrum of Q is on the circle $|\lambda|=1$, it follows that $\lambda=-1$ occurs in the point spectrum of Q with a multiplicity, say $l(\geqq 0)$, for which the difference $k-l$ must become even. As explained above, this proves the first of the two assertions of (II).

16. In order to prove the remaining assertion of (II), suppose that -1 is in the essential spectrum of an orthogonal matrix O. The assertion to be proved is that both rotations and reflections must then cluster at the point O of Ω.

Since this assertion remains unaltered if O is replaced by a matrix orthogonally equivalent to O, it can be assumed that O is in one of its three normal forms (16), with (17)-(18). But -1 then is in the essential spectrum of O either because it is in the spectrum of C or because -1 is either a cluster point or a point of infinite multiplicity in the point spectrum of P (the latter P can belong either to the second or to the third of the three cases in (16)).

---

[2] By eigenvalues are meant points in the point spectrum and, when estimating their number, they are meant to be enumerated so as to take into account their multiplicities.

An orthogonal matrix will be said to be of type (*) if -1 occurs in its point spectrum with a finite (possibly zero) multiplicity. Let O_1 and O_2 be of type (*); then so also is $O_1 \dotplus O_2$, and (I) implies the following fact: If O_1 is a rotation (reflection), then the orthogonal matrix $O_1 \dotplus O_2$ is a rotation (reflection) if and only if O_2 is a rotation. Hence it is easy to see from (16) and (17)-(18) that it is sufficient to show that O is a cluster point both of rotations and of reflections, *of type* (*), in the following three particular cases:

(i) O has no point spectrum and -1 is in the (continuous) spectrum.

(ii) O has no continuous spectrum and -1 is a cluster point of the (point) spectrum but is not in the point spectrum.

(iii) $O = -(1 \dotplus 1 \dotplus \cdots)$, i. e., $O = -I$.

(Moreover, it is seen that a portion of the assertion of (II) can be improved as follows: If -1 is in the essential spectrum of an orthogonal matrix O, then O is a cluster point both of rotations and reflections of type (*).)

Case (iii) is straightforward. In fact, if $0 < \epsilon < \pi$, and if $B(\phi)$ denotes again the binary rotation by the angle ϕ, then it is seen from (I) that

$$P_\epsilon = B(\pi - \epsilon/1) \dotplus \beta(\pi - \epsilon/2) + B(\pi - \epsilon/3) \dotplus \cdots$$

is a rotation, and that $-1 \dotplus P_\epsilon$ is a reflection. On the other hand, since $1/1^2 + 1/2^2 + \cdots < \infty$, and since $-B(\pi)$ is the binary unit matrix, it is clear that both the rotation P_ϵ and the reflection $-1 \dotplus P_\epsilon$ are of type (*) and tend, as $\epsilon \to 0$, to $-I$, which is the O of case (iii).

Case (ii) can be disposed of similarly. First, it is seen from (17) that a normal form of O in the case (ii) is

$$(24) \qquad O = B(\phi_1) \dotplus B(\phi_2) \dotplus \cdots \text{ or } O = 1 \dotplus B(\phi_1) \dotplus B(\phi_2) \dotplus \cdots,$$

where the angles $\phi_1, \phi_2, \cdots$ satisfy the following pair of conditions: $|\phi_m| < \pi$ holds for every m, and $\liminf |\pi - \phi_m| = 0$ as $m \to \infty$. In view of the *first* of these conditions, (I) and (24) show that O is a rotation, and that it remains a rotation (and of type (*)) if one term of (24), say the first, is changed from $B(\phi_1)$ to $B(\phi_1 \pm \epsilon)$, provided that $0 < \epsilon < \pi - |\phi_1|$. But if O_ϵ denotes the matrix which thus results from (24), then $O_\epsilon \to O$ as $\epsilon \to 0$, and so O is a cluster point of rotations of type (*). In order to prove that O is a cluster point of reflections of type (*) as well, use must be made of the *second* of the conditions, which is $\liminf |\pi - \phi_m| = 0$ as $m \to \infty$.

Since the latter condition implies that $\phi'_m \to \pi$ holds for a suitable infinite subsequence $\phi'_1, \phi'_2, \cdots$, of $\phi_1, \phi_2, \cdots$, it is clear that, with reference to every $\epsilon > 0$, it is possible to write (24) in the form

$$(25) \qquad O = O_\epsilon \dotplus (-I + A_\epsilon),$$

where the plus sign occurring in $-I + A_\epsilon$ refers to *ordinary* matrix addition, $-I$ is $B(\pi) \dotplus B(\pi) \dotplus \cdots$, and the infinite matrix A_ϵ represents a correction term, with $| A_\epsilon | < \epsilon$. It follows from (I), and from the fact that O is in case (ii), that O_ϵ is a rotation not containing -1 in its point spectrum. It follows therefore from the remarks preceding the statements of cases (i)-(iii) that O will be proved to be a cluster point of reflections of type (*) if it is ascertained that, for every $\epsilon > 0$, the inequality $| (-I + A_\epsilon) - R | < 2\epsilon$ can be satisfied by a certain reflection $R = R(\epsilon)$. But since $| A_\epsilon | < \epsilon$, the existence of such an $R = R(\epsilon)$ follows if $| -I - R_m | \to 0$, as $m \to \infty$, holds for some sequence of reflections $R_1, R_2, \cdots$ of type (*). Since the existence of such reflections R_m was proved in the treatment of case (iii), the treatment of case (ii) is now complete.

17. The treatment of the remaining case, (i), will now be reduced to the following

LEMMA. *If -1 is in the spectrum, but not in the point spectrum, of an orthogonal matrix O, then there exists a sequence of orthogonal matrices $O_1, O_2, \cdots$ satisfying $| O - O_m | \to 0$, as $m \to \infty$, and having the property that -1 is a cluster point of the point spectrum, but is not in the point spectrum, of O_m, where $m = 1, 2, \cdots$.*

Needless to say, the matrices O and $O_1, O_2, \cdots$ are rotations by necessity: cf. (I).

It is clear that if an O is in case (i), then it satisfies the assumptions of the Lemma (but the converse is not true). If a matrix O_m of the above sequence does not possess any continuous spectrum, then it is in case (ii), so that both rotations and reflections (of type (*)), say R_k, must cluster at O_m. The last assertion remains true even if O_m does possess a continuous spectrum. In fact, if $O_m = C \dotplus P_\infty$ (in accordance with (16); note that -1 is a cluster point of the point spectrum), then P_∞ is in case (ii), so that both rotations and reflections R_k of type (*) cluster at P_∞; hence both rotations and reflections $C \dotplus R_k$ of this same type cluster at O_m; cf. the remarks preceding the statements of (i)-(iii) in § 16. It follows therefore from the Lemma that every O of the case (i) must be a cluster point (on Ω) of rotations and reflections of type (*). Accordingly, the proof of (II) will be complete if the Lemma is proved.

18. To this end, let O be a matrix satisfying the assumptions of the Lemma. Then 1 is in the essential spectrum, but not in the point spectrum, of the orthogonal matrix $-O$. Hence, if $E(\lambda)$ denotes the spectral matrix, satisfying (21a)-(22b), of O, then $E(\lambda)$ is not constant on any interval $0<\lambda<\eta$ and is continuous at $\lambda=0$. In view of the parenthetical remark in (22a), the latter property means that

$$E(+0)=E(0)=\| 0 \|. \tag{26}$$

On the other hand, the non-constancy of $E(\lambda)$ on any interval $0<\lambda<\eta$ means that, with reference to every positive ϵ which is less than π, it is possible to choose a sequence of values $\lambda=\lambda_n=\lambda_n(\epsilon)$ satisfying

$$\epsilon=\lambda_0>\lambda_1>\cdots>\lambda_n\rightarrow 0, \text{ where } n\rightarrow\infty, \tag{27}$$

and

$$E(\lambda_n)\neq E(\lambda_{n+1}), \text{ where } n=0,1,2,\cdots. \tag{28}$$

Disregarding an enumerable set of λ-values, one may assume that $\lambda_0, \lambda_1, \lambda_2, \cdots$ are continuity points of $\bar{E}(\lambda)+E(2\pi-\lambda)$.

With reference to the sequence (27), where $\lambda_n=\lambda_n(\epsilon)$, define for $-\infty<\lambda<\infty$ a matrix function $F(\lambda)=F_\epsilon(\lambda)$ as follows:

$$F(\lambda)=E(\lambda_n) \text{ if } \lambda_{n+1}<\lambda\leqq\lambda_n \text{ and } F(2\pi-\lambda)=E(2\pi-\lambda_n) \text{ if } \lambda_{n+1}\leqq\lambda<\lambda_n \tag{29}$$

and

$$F(\lambda)=E(\lambda) \text{ if } \epsilon<\lambda\leqq 2\pi-\epsilon \qquad (\epsilon=\lambda_0), \tag{30}$$

finally $F(\lambda)=\| 0 \|$ if $\lambda\leqq 0$ and $F(\lambda)=I$ if $\lambda\geqq 2\pi$. Then it is seen from (26) and (21a)-(22b), where $E(\lambda)$ is the spectral matrix of a Hermitian matrix, that

$$F(+0)=F(0)=\| 0 \|, \qquad F(2\pi-0)=F(2\pi)=I \tag{31}$$

and that $F(\lambda)$ is the spectral matrix of the Hermitian matrix

$$H=\int_{-\infty}^{\infty}\lambda dF(\lambda)=\int_{+0}^{2\pi-0}\lambda dF(\lambda), \tag{32}$$

and therefore that of the unitary matrix e^{iH}.

Clearly, $F(\lambda)$ is a step-function on the outside of the interval $\epsilon<\lambda<2\pi-\epsilon$, with jumps possible, for $\lambda\leqq\epsilon$, only at the λ-values (27), and which, in view of (28), take place at each of the values (27), clustering at 0. Hence e^{iH} has no continuous spectrum near the value $e^{i0}=1$, and its point spectrum clusters at the value $e^{i0} = 1$, which value is not in the point spectrum of e^{iH}.

In addition, since conditions (21a)-(22b) are satisfied by the given $E(\lambda)$, it is seen from (29), (30) and (31) and the properties of the sequence $\lambda_0, \lambda_1, \lambda_2, \cdots$ that these conditions are satisfied if E is replaced by F. But the satisfaction of (22a)-(22b) by $E = F$ means that the unitary matrix having the spectral matrix $F(\lambda)$ is real, i. e., orthogonal.

Accordingly, e^{iH} is an orthogonal matrix which has no continuous spectrum near the value 1, and which contains 1 is in its essential spectrum but not in its point spectrum.

19. By construction, H is here a function, H_ϵ, of ϵ, where $0 < \epsilon < \pi$. It will now be shown that

$$(33) \qquad | e^{iH} + O | \to 0 \text{ as } \epsilon \to 0 \text{ in } H = H_\epsilon.$$

First, since $E(\lambda)$ and $F(\lambda)$ are spectral matrices,

$$(34) \qquad | \int_0^\epsilon \lambda dE(\lambda) - \int_0^\epsilon \lambda dF(\lambda) | \leqq \epsilon + \epsilon.$$

In virtue of (26), (31) and (21a) (the last being valid for E itself and for $E = F$),

$$\exp \int_0^\epsilon i\lambda dE(\lambda) = I - E(\epsilon) + \int_0^\epsilon e^{i\lambda} dE(\lambda)$$

holds for E itself and for $E = F$. It is seen from (27). (29) and (34) that, as $\epsilon \to 0$,

$$\int_0^\epsilon e^{i\lambda} dE(\lambda) = \int_0^\epsilon e^{i\lambda} dF(\lambda) = o(1).$$

Disregarding an enumerable set of ϵ-values, choose ϵ so that $\lambda = \epsilon$ is a continuity point of $E(\lambda)$ and of $E(2\pi - \lambda)$ (and hence of $F(\lambda)$ and of $F(2\pi - \lambda)$; cf. (29) and (30)). Then, by the last formula line and (22b), valid for $E = F$ as well as for E itself, the relation

$$\int_{2\pi-\epsilon}^{2\pi} e^{i\lambda} dE(\lambda) - \int_{2\pi-\epsilon}^{2\pi} e^{i\lambda} dF(\lambda) = o(1), \qquad \epsilon \to 0,$$

holds; cf. (22b), valid for E itself and for $E = F$, and note that the present

$E(\lambda)$ and $F(\lambda)$ are continuous at $\lambda = 0$ as well as at $\lambda = 2\pi$. Finally, from (30),

$$\int_{\epsilon+0}^{2\pi-\epsilon-0} e^{i\lambda} dE(\lambda) - \int_{\epsilon+0}^{2\pi-\epsilon-0} e^{i\lambda} dF(\lambda) \doteq 0.$$

Addition of the last three formula lines shows that

$$\int_0^{2\pi} e^{i\lambda} dE(\lambda) - \int_0^{2\pi} e^{i\lambda} dF(\lambda) = o(1) \tag{35}$$

as $\epsilon \to 0$, where $F(\lambda) = F_\epsilon(\lambda)$. But the difference on the left of (35) is identical with $-O - e^{iH}$, since $E(\lambda)$ and $F(\lambda)$ have been defined as the spectral matrices of $-O$ and e^{iH}, respectively. Consequently, (35) proves (33).

20. If $m = 1, 2, \cdots$, let O_m denote the matrix $-e^{iH}$ which results if ϵ in $H = H_\epsilon$ is chosen to be $\epsilon_m (> 0)$, where $\epsilon_m \to 0$ as $m \to \infty$. Then $| O_m - O | \to 0$, by (33). On the other hand, the last sentence of § 18 shows that every O_m has all the properties claimed in the Lemma of § 17. Thus the proof of the Lemma (and hence that of (II)) is complete.

The "Boundary" $[\mathbf{\Omega}]$ of the $\mathbf{\Omega}$-Space.

21. If $E(\lambda)$, where $0 \leqq \lambda \leqq 2\pi$, is the spectral matrix of an orthogonal matrix O, define the number $\gamma = \gamma(O)$, where $0 \leqq \gamma \leqq \pi$, as follows: According as $\lambda = \pi$ is not or is a point of constancy of $E(\lambda)$ (i. e., according as -1 is or is not in the spectrum of O), let γ denote 0 or the largest number ($\leqq \pi$) having the property that $E(\lambda) = E(\pi)$ whenever $| \lambda - \pi | < \gamma$. Then

$$| I - O | = 2 \sin \tfrac{1}{2}(\pi - \gamma) \tag{36}$$

holds in either case. In fact, since

$$I - O = \int_0^{2\pi} (1 - e^{i\lambda}) dE(\lambda),$$

it is readily seen from the definition of $\gamma = \gamma(O)$ that $| 1 - e^{i(\pi-\gamma)} |$ is the least upper bound of the length of the vector $(I - O)x$ when x varies over all vectors of unit length. But this evaluation of l. u. b. $|(I - O)x |/| x |$ is equivalent to (36).

It is seen from the definition, (2), of the "boundary," $[\Omega]$, of Ω that the particular case $\gamma = 0$ of (36) can be formulated as follows:

(i) *An orthogonal matrix, O, represents a point on the "boundary," $[\Omega]$, if and only if -1 is in the spectrum of O.*

This criterion, (i), leads to the following corollary:

(ii) *The "boundary," $[\Omega]$, contains all reflections and some rotations.*

In fact, (I) implies that -1 is in the spectrum of O if, but not only if, O is a reflection. Hence (ii) follows from (i). Clearly, (14) is just a restatement of (ii).

22. Let F denote the diagonal matrix $-1 + D$, where D is the orthogonal matrix defined at the beginning of § 13. Then the essential spectrum of either of the matrices D, F, being the single value 1, does not contain the value -1. Furthermore, (I) implies that F is a rotation and that D is a reflection. It follows therefore from (II) that the rotation F is not in the closure of all reflections of Ω and that the reflection D is not in the closure of all rotations of Ω. But (i) and the definitions of D and F show that both D and F are in $[\Omega]$. Since $[\Omega]$ is a subset of Ω, it follows that $[\Omega]$ contains a reflection $(= D)$ which is not in the closure of the rotations contained in $[\Omega]$. It is also seen that $[\Omega]$ contains a rotation $(= F)$ which is not in the closure of the reflections contained in $[\Omega]$, since, according to (ii), all reflections are in $[\Omega]$. Accordingly, the situation is as follows:

(iii) *Neither the set of all reflections nor the set of all rotations contained in $[\Omega]$ is dense on $[\Omega]$.*

It was also seen that $[\Omega]$ contains a point $(= D)$ which fails to be in the closure of *all* rotations of Ω. Since $\Omega - [\Omega]$ is a subset of Ω, it follows that the names "boundary of Ω" and "interior of Ω" for the two sets (2) are not justifiable in terms of the topology of Ω:

(iv) *The "boundary," $[\Omega]$, of Ω is not a subset of the closure of the "interior," $\Omega - [\Omega]$, of Ω (i.e., Ω is not the closure of $\Omega - [\Omega]$).*

What is true (but trivial) is the following fact:

(v) *The "boundary," $[\Omega]$, is a closed set (i.e., the "interior," $\Omega - [\Omega]$, is an open set) on Ω.*

For, if O is any point of Ω for which $|I - O|$ is less than 2, say $2 - \epsilon$, then, since Ω is a metric space, $|I - Q| < 2 - \frac{1}{2}\epsilon (< 2)$ holds whenever

$| Q - O | < \frac{1}{2}\epsilon$. In view of (2), this means that $\Omega - [\Omega]$ is an open set. Hence (v) follows by complementation.

23. By a dualization of the content of (iii), the following facts will now be proved:

(vi) *Neither the set of the reflections nor the set of the rotations contained in* $[\Omega]$ *is an open set (hence, neither of these sets is a closed set) on* Ω.

In view of (ii), those assertions of (vi) which concern the reflections are contained in the corresponding assertions of (*), § 2, proved in § 13. Hence, in order to prove (vi), it is sufficient to show that the product $[\Omega]\Omega_0$, representing that portion of the set, Ω_0, of all rotations which is on the "boundary" of Ω, is neither an open set nor a closed set (on Ω). In other words, it is sufficient to show that the set $[\Omega]\Omega_0$ contains a point, say Q, and that the set $\Omega - \Omega_0$ (which is identical with the product of $[\Omega]$ and $\Omega - \Omega_0$) contains another point, say R, such that Q is in the closure of $\Omega - \Omega_0$ and R is in the closure of $[\Omega]\Omega_0$. The existence of Q is assured by (II) if Q is chosen to be any rotation with -1 in its essential spectrum.

In order to guarantee the existence of R one can proceed as follows: Let A denote a rotation for which -1 is in the essential spectrum but is not the point spectrum. Then $A_1 = -1 \dot{+} A$ is a reflection by (I), while (II) shows there exists a sequence of reflections, B_n, such that $B_n \to A$, hence $-1 \dot{+} B_n \to A_1$. Another application of (I) shows that the matrices $-1 \dot{+} B_n$ are rotations (which clearly belong to $[\Omega]$). The point R may now be chosen to be A_1 and the proof of (vi) is complete.

24. In view of the parenthetical restatement in (iv), there arises the question concerning a characterization of those orthogonal matrices which are in the closure of $\Omega - [\Omega]$. The situation will be described by

(III) *A point, O, of Ω is not in the closure of $\Omega - [\Omega]$ if and only if O is in the interior of $\Omega - \Omega_0$.*

In contrast to (iv), the definition of an interior point is now based on the topology of Ω, as determined by the "strong" $| O_1 - O_2 |$-metric of § 1. In this terminology, the Corollary of (II) can of course be restated as follows:

(II bis) *Whether a point, O, of Ω be in Ω_0 or in $\Omega - \Omega_0$, it is an interior point of Ω_0 or of $\Omega - \Omega_0$ if and only if -1 is not in the essential spectrum of the orthogonal matrix represented by the point O.*

This formulation, (II bis), of (II) is adjusted to the following proof of (III).

25. Consider first that assertion of (III) in which a given O is supposed to be a point in the interior of $\Omega - \Omega_0$. Thus, in a sufficiently small vicinity of O (with reference to the "strong" $| O_1 - O_2 |$-metric), there exist no rotations. Since the set $\Omega - [\Omega]$ consists only of rotations, the proof of the first assertion of (III) is complete.

The remaining assertion of (III) is that if a point, O, of Ω is not in the interior of $\Omega - \Omega_0$, then

$$O \text{ is in the closure of } \Omega - [\Omega]. \tag{37}$$

But an O is not in the interior of $\Omega - \Omega_0$ either because (a) the point O is in Ω_0 or because (b) the point O is in $\Omega - \Omega_0$ and in the closure of Ω_0. It is however clear that if (37) is granted to be true in case (a), then (37) must be true in case (b) also. Hence it is sufficient to prove (37) in case (a), i. e., under the assumption that O is a rotation. Then (I) shows that, if O is assumed to be given in its appropriate normal form (16)-(18), the P_n occurring in the last of the three possibilities (16) cannot be of type (18) and must therefore be of type (17). On the other hand, it is readily seen that (37) will be proved for the third of the three cases (16) if it is proved for the second and for the first.[3] Consequently, if the trivial term 1 mentioned in (17) is omitted, it is sufficient to prove (37) under the assumption that the rotation O is either of the form

$$O = B(\phi_1) \dot{+} B(\phi_2) \dot{+} \cdots, \text{ where } | \phi_m - \pi | \leqq \pi, \tag{38}$$

or of the form $O = C$, where $B(\phi)$ is a binary rotation and C has no point spectrum.

Consider first the case $O \neq C$. Then it is clear from (38) that, corresponding to every positive ϵ which is less than π, it is possible to choose a sequence of ϕ-values $\phi_1{}^\epsilon, \phi_2{}^\epsilon, \cdots$ satisfying the following two conditions: $\epsilon \leqq | \phi_m{}^\epsilon - \pi | \leqq \pi$ holds for every m and $| O - O^\epsilon | \to 0$ holds as $\epsilon \to 0$, where O^ϵ denotes the orthogonal matrix $B(\phi_1{}^\epsilon) \dot{+} B(\phi_2{}^\epsilon) \dot{+} \cdots$. Because of the second of these conditions, (37) will follow if it is ascertained that the point O^ϵ of Ω is in $\Omega - [\Omega]$, i. e., that O^ϵ is not in $[\Omega]$. But the first of the two conditions shows that every value occurring in the spectrum of O^ϵ differs from $-1 = e^{i\pi}$ by not less than ϵ (in angular distance). Hence, -1 is not in the spectrum of O^ϵ. It follows therefore from (i), § 21, that O^ϵ is not in $[\Omega]$. This proves (37) for the case (38).

---

[3] This is clear when n is ∞ in the third case of (16); on the other hand, if n is finite (and hence, by (I), even), the proof will be clear from the consideration below, the treatment for P_n being essentially similar to that for P_∞.

26. Similarly, (37) will be proved for the remaining case, where $O = C$, if it is shown that there belongs to every $\epsilon > 0$ an orthogonal matrix, say O_ϵ, having the following properties: The value -1 is not in the spectrum of O_ϵ and $| C - O_\epsilon | \to 0$ holds as $\epsilon \to 0$. But the construction of such an O_ϵ hardly differs from the construction of O^ϵ in the case of (38).

First, if $E(\lambda)$, where $0 \leqq \lambda \leqq 2\pi$, denotes the spectral matrix of $-C$, then, since C has no point spectrum, $E(\lambda)$ is continuous throughout. If $0 < \epsilon < \pi$, define $F(\lambda)$ by $F(\lambda) = \| 0 \|$, $E(\lambda)$ or I according as $0 \leqq \lambda \leqq \epsilon$, $\epsilon < \lambda \leqq 2\pi - \epsilon$ or $2\pi - \epsilon < \lambda \leqq 2\pi$. Thus the conditions (21a)-(22b), which are satisfied by the given E, become satisfied by F. It is seen that (34) is also valid and the procedure is now similar to that of § 19. It is clear that $-e^{iH}$ does not have -1 in its spectrum and that $| C + e^{iH} | \to 0$ as $\epsilon \to 0$. Thus O_ϵ can be defined to be $-e^{iH}$ and the proof of (III) is complete.

Arcwise Connections.

27. If Λ is a subset of Ω and if P, Q is a pair of points in Λ, let the symbol $P(\Lambda)Q$ denote that either $P = Q$ or, if $P \neq Q$, there exists in Ω a continuous path $O = O(t)$, $a \leqq t \leqq b$ $(\geqq a)$, having the following properties: $O(t)$ is a point of Λ (for every t) and $O(a) = P$, $O(b) = Q$. Thus $Q(\Lambda)P$ implies that $P(\Lambda)Q$ and that $Q(\Lambda)\mathrm{R}$ whenever $P(\Lambda)R$.

If $P(\Lambda)Q$ holds for every pair of points P, Q contained in Λ, then Λ is called arcwise connected. This is the meaning of the six assertions (9)-(13), (15) of the introduction, which have not been used thus far. The purpose of this chapter is to prove the last four of those six assertions. The first two of them have already been proved (cf. § 4).

The proofs will involve the circumstance that if Λ is any of the four sets in question, then Λ is an invariant subset of Ω (simply because the same is true of both $\Lambda = \Omega_0$ and $\Lambda = [\Omega]$). In other words, all the subsets Λ of Ω which are to be considered have the property

$$(39) \qquad R\Lambda R^{-1} = \Lambda,$$

where R is *any* orthogonal matrix (so that R need not be in Λ). It is clear from (9) that $R(\Omega)I$. hence $O(\Lambda)ROR^{-1}$, so that, in order to prove that an O-set Λ satisfying (39) is arcwise connected, it is sufficient to exhibit in Λ a single point O_0 having the property that $ROR^{-1}(\Lambda)O_0$ holds for every point O of Λ and some point R of Ω (where R, in contrast to the fixed O_0, is a function of O). In particular, it is allowed to assume that O is given in an orthogonal normal form, (16)-(18), even though such a form of a given matrix of Λ cannot in general be attained by using orthogonal transformations contained in Λ.

28. The proofs will be deduced from the following theorem:

(IV) *Let R be an orthogonal matrix, $E(\lambda)$ the spectral matrix of $-R$, finally t a parameter on the range $0 \leqq t \leqq 1$. Then*

(α) *if -1 is not in the point spectrum of R, the matrix defined by*

$$R_t = \int_0^{2\pi} e^{it(\lambda-\pi)} dE(\lambda); \qquad 0 \leqq t \leqq 1, \qquad (R_1 = R_0, R_0 = I), \tag{40}$$

is orthogonal, and -1 is not in its point spectrum (for any t);

(β) *if -1 is not in the spectrum of R, it is not in the spectrum of R_t;*

(γ) *under the assumption made in (α),*

$$R_u R_v = R_{u+v}. \tag{41}$$

Since the assumption in (α) means (and therefore the assumption in (β) implies) that 1 is not in the point spectrum of $-R$, both $\lambda = 0$ and $\lambda = 2\pi$ are continuity points of the spectral matrix, $E(\lambda)$, of $-R$, and so

$$\int_0^{2\pi} = \int_{+0}^{2\pi-0} \tag{42}$$

in (40). (Actually, $\lambda = 2\pi$ is always a continuity point for any spectral matrix $E(\lambda)$ under the normalization (21a).) It is readily verified from (21a)-(22b) that the matrix (40), where $0 \leqq t \leqq 1$, is real. But (40) surely is a unitary matrix (in fact, $E(\lambda)$ is a spectral matrix for $0 \leqq \lambda \leqq 2\pi$). Consequently, (40) is an orthogonal matrix, as claimed by the first part of (α).

In order to prove the second claim of (α), suppose, if possible, that -1 is in the point spectrum of R_t (for some t). This means that (for that t) there exists a unit vector x satisfying $(I + R_t)x = 0$. But it is seen from (40) and (42) that

$$|(I + R_t)x|^2 = \int_{+0}^{2\pi-0} |1 + e^{it(\lambda-\pi)}|^2 d\,|E(\lambda)x|^2 \tag{43}$$

holds for every unit vector x, not only for those satisfying the assumption $(I + R_t)x = 0$. Hence, the latter implies the vanishing of the integral (43). But

$$(44)\qquad \int_{+0}^{2\pi-0} d\,|\,E(\lambda)x\,|^2 = \int_0^{2\pi} d\,|\,E(\lambda)x\,|^2 = |\,x\,|^2$$

is an identity in x. On the other hand, it is seen from the assumption $0 \leqq t \leqq 1$ that the function integrated in (43) is positive at every point λ of the (open) interval of integration. Consequently, the vanishing of the integral (43) implies that $x = 0$. Since this contradicts the assumption $|\,x\,| = 1$, the proof of (α) is complete.

In order to prove (β), suppose that -1 is not in the spectrum of R. This means that $\lambda = 0$ and $\lambda = 2\pi$ are points of constancy (rather than, as in the more general case of (α), just points of continuity) of $E(\lambda)$, i. e., that (42) can be improved to

$$(45)\qquad \int_0^{2\pi} = \int_\mu^{2\pi-\mu}, \text{ where } 0 < \mu < \pi.$$

But $0 < \mu < \pi$ and $0 \leqq t \leqq 1$ imply that the function integrated in (43) has a positive minimum on the interval $\mu \leqq \lambda \leqq 2\pi - \mu$. It follows therefore from (43), (44) and (45) that $|\,(I + R_t)x\,|^2 \geqq \text{const.}\,|\,x\,|^2$, where the const. is positive and independent of x. In other words, the quadratic form $|\,(I + R_t)x\,|^2$ is positive definite. But a classical criterion of Toeplitz (cf., e. g., [4], p. 138) implies that if A is a real, bounded, normal matrix, then A^{-1} exists (as a unique, bounded reciprocal) if and only if the quadratic form $|\,Ax\,|^2$ is positive definite. Consequently, $I + R_t$ has a bounded reciprocal matrix. This proves part (β) of (IV).

Finally, in order to prove (γ), it is sufficient to observe that, since $E(\lambda)$ is a spectral matrix on $0 \leqq \lambda \leqq 2\pi$, the (unitary) matrix

$$W_t = \int_0^{2\pi} e^{i\lambda t} dE(\lambda) \quad \text{satisfies} \quad W_{u+v} = \int_0^{2\pi} e^{i\lambda u} e^{i\lambda v} dE(\lambda) = W_u W_v,$$

and that $R_t = e^{-i\pi t} W_t$, by (40).

29. *Proof of* (11). Let I^{2h-1}, where $h = 1, 2, \cdots$, denote the $(2h-1)$-rowed unit matrix. Then (I) shows that an orthogonal matrix is a reflection if and only if it is orthogonally equivalent to a matrix of the form $-I^{2h-1} \dotplus R$, where R denotes a rotation not containing -1 in its point spectrum. Since $-1 \dotplus I = -I^1 \dotplus I$ is a reflection, it follows from the remarks made at the end of § 27 that (11) is equivalent to the following statement:

$$(46)\qquad (-I^{2h-1} \dotplus R)\,(\Lambda)\,(-I^1 \dotplus I), \text{ where } \Lambda = \Omega - \Omega_0.$$

Here R denotes any rotation not containing -1 in its point spectrum. But part (a) of (IV) shows that any such R can be connected to I by a continuous path in such a way that -1 is not contained in the point spectrum of any matrix, R_t, representing a point of the path. Since this means that every point of the path R_t is a rotation, every point of the path $-I^{2h-1} \dot{+} R_t$ is a reflection, i. e., a point of $\Omega - \Omega_0$. But the end points, $t = 0$ and $t = 1$, of the latter path are $-I^{2h-1} \dot{+} I$ and $-I^{2h-1} \dot{+} R$. Consequently,

$$(-I^{2h-1} \dot{+} R)(\Omega - \Omega_0)(-I^{2h-1} \dot{+} I).$$

Hence, in order to prove (46), it is sufficient to ascertain the truth of the following relation:

$$(47) \qquad (-I^{2h-1} \dot{+} I)(\Lambda)(-I^1 \dot{+} I), \text{ where } \Lambda = \Omega - \Omega_0.$$

In order to prove (47), consider the matrix

$$(48) \qquad -I^1 \dot{+} B(\pi t) \dot{+} \cdot \cdot \cdot \dot{+} B(\pi t) \dot{+} I,$$

in which the number of B-terms is chosen to be $h - 1$ ($< \infty$, possibly 0). Since $B(\phi)$ denotes the binary matrix representing a rotation by the angle ϕ, it is clear that (48) is a reflection (i. e., a point of $\Omega - \Omega_0$) for every t, and that (48) becomes $-I^1 \dot{+} I$ at $t = 0$ and $-I^{2h-1} \dot{+} I$ at $t = 1$. This proves (47). Hence the proof of (11) is now complete.

30. *Proof of* (12). Let (15), to be proved below, be granted. Then it is clear from (11) and (14) that, in order to prove (12), it is sufficient to exhibit on $[\Omega]$ *one* rotation and *one* reflection, say O_1 and O_2, which can be joined by a continuous path contained in $[\Omega]$. But (i) in § 21 shows that both matrices $-1 \dot{+} (\pm I)$ are in $[\Omega]$ and, according to (I), only one of them is a reflection. Consequently, (12) will follow if it is ascertained that

$$(49) \qquad O^-([\Omega])O^+, \text{ where } O^{\pm} = -1 \dot{+} (\pm I).$$

But it is seen from (i) in § 21 that the truth of (49) is equivalent to the truth of

$$(50) \qquad -I(\Omega)I.$$

Finally, (50) is an obvious consequence of (9). In fact, since both matrices $\pm I$ are rotations, (10) implies that

$$(50 \text{ bis}) \qquad -I(\Omega_0)I,$$

which is more than (50).

31. *Proof of* (13). Criterion (i) of § 21 shows that a point O of Ω is in $\Omega - [\Omega]$ if and only if -1 is not in the spectrum of O. This implies that I is in $\Omega - [\Omega]$ and that (13) will be proved if it is shown that

$$O(\Omega - [\Omega])I \tag{51}$$

holds whenever O is an orthogonal matrix not containing -1 in its spectrum.

In order to prove (51) for every such O, apply part (β) of (IV) to $R = O$. This supplies the existence of a continuous path $R_t (0 \leqq t \leqq 1)$ which begins at the point I, ends at the point $R = O$, and is such that no matrix R_t, representing an arbitrary point of the path, will have -1 in its spectrum. In view of (i), § 21, this means that every point of the path will be in $\Omega - [\Omega]$. Since the path connects I to O, the truth of (51) follows.

32. *Proof of* (15). According to (i), § 21, and (I), an O is in the intersection of $[\Omega]$ and Ω_0 if and only if -1 is in the spectrum of O and occurs in the point spectrum of O with a multiplicity which is either ∞ or even ($\geqq 0$). Clearly, O is in $[\Omega]\Omega_0$ if and only if it is orthogonally equivalent to a matrix of the form

$$O = R \quad \text{or} \quad O = I^j \dotplus R, \tag{52}$$

where I^j is the j-rowed unit matrix ($1 \leqq j \leqq \infty$) and R is a rotation which has -1 in its spectrum and does not have $+1$ in its point spectrum. (In the first case of (52), R is, of course, infinite; in the second case, R may be either finite or infinite.) It is sufficient to show that O can be connected by a path in $[\Omega]\Omega_0$ to a fixed matrix of this space, say to $-I$.

Let

$$R = \int_{+0}^{2\pi-0} e^{i\lambda} dE(\lambda)$$

denote the spectral resolution of R, and define R_t, for $0 \leqq t \leqq 1$, by

$$R_t = -\int_{+0}^{2\pi-0} e^{it(\lambda-\pi)} dE(\lambda).$$

Then the matrices R_t are orthogonal for all t and join $R_0 = -I^{2k}$ to $R_1 = R$; cf. (IV). It is easy to see that -1 is, for $0 < t \leqq 1$, in the spectrum of R_t, and that the multiplicity of -1 in the point spectrum of R_t is identical with that of -1 in the point spectrum of R. (In fact, if $n \to \infty$, then $(R + I)x_n \to \infty$ holds for a sequence of unit vectors x_n if and only if

$(R_t+I)x_n \to 0$ holds for every fixed t on $0 < t \leqq 1$; while $(R+I)x=0$ holds for a vector $x \neq 0$ if and only if $(R_t+I)x=0$ holds for every fixed t on $0 < t \leqq 1$; cf. the proof of (IV) in § 28.) Consequently, it is clear from (52) that $O([\Omega]\Omega_0)N$, where $N=-I$ or $N=M$, with

$$M = I^j \dotplus (-I^k), \text{ where } 1 \leqq j \leqq \infty \text{ and } 1 \leqq k \leqq \infty.$$

Since the proof of (15) is complete if $N=-I$, it remains only to show that

$$M([\Omega]\Omega_0)(-I). \tag{53}$$

If $j=\infty$, then $I^j=I$, hence (53) follows, since I can be connected to $-I$ in Ω_0. If $j<\infty$, then $M=I^j \dotplus (-I) \dotplus (-I)$. But $I^j \dotplus (-I)$ is a rotation and can therefore be joined to $-I$ in Ω_0, so that (53) follows as before. This completes the proof of (15).

33. Assertions (A) and (B) of § 5 will now be proved. In order to prove (A), it is sufficient to note that $Q(t)=B(\pi t) \dotplus I$, where $0 \leqq t \leqq 1$ and where $B(\phi)$ is the binary rotation defined in § 7, is a continuous path satisfying the following conditions: $Q(1)$ is in $[\Omega]$, $Q(t)$ is in $\Omega - [\Omega]$ when $0 \leqq t < 1$, finally $I = Q(0)$. Assertion (B) is a consequence of (II); one need only choose the O of (B) to be a reflection which does not have -1 in its essential spectrum.

APPENDIX. On Bounded Matrices.

34. The following considerations concern themselves with facts which correspond to those of §21-§26 and (9)-(15) if Ω, the space of all (real) orthogonal matrices, is replaced by the larger space, say Θ, of all bounded real matrices, A. It turns out that a formal analogy results if what corresponds to the "boundary," $[\Omega]$, of Ω is taken to be the set, say $[\Theta]$, of those matrices A which fail to have a (unique) bounded reciprocal matrix, A^{-1}. Thus $\Theta - [\Theta]$ and $[\Theta]$ consist of all non-singular and of all singular matrices which represent a linear mapping of the real Hilbert space, $\mathfrak{R}$, on the whole of $\mathfrak{R}$ or on a proper subset of $\mathfrak{R}$, respectively. It is understood that Θ is meant to be the metric space on which the distance, $|A_1 - A_2|$, is defined as at the beginning of § 1.

35. What corresponds to (v) in § 22 is the following fact (which is true for finite matrices also):

(*a*) *The set* $[\Theta]$ *is closed, i. e., the set* $\Theta - [\Theta]$ *is open (on* Θ*).*

In order to prove this assertion, (α), it is sufficient to note that Θ is a complete space and that, if A has a unique, bounded reciprocal A^{-1} (i. e., if A is in $\Theta - [\Theta]$), then the partial sums of both series

$$\sum_{n=0}^{\infty} A^{-1}(-A^{-1}X)^n, \qquad \sum_{n=0}^{\infty} (-XA^{-1})^n A^{-1},$$

the formal Liouville series for $(A+X)^{-1}$, form convergent sequences on Θ whenever $|X|$ is small enough (smaller than $1/|A^{-1}|$).

Clearly, (α) implies the first of the following two assertions:

(β) *Neither* $[\Theta]$ *nor* $\Theta - [\Theta]$ *is dense on* Θ.

The second assertion of (β) is false for finite matrices and represents the analogue of (iv), § 22. It will be proved by showing that there exist on Θ points A which are not in the closure of $\Theta - [\Theta]$. In fact, it will be proved that an A satisfies this condition if the spectra of AA' and $A'A$ are not identical,

$$\text{(54)} \qquad \mathrm{sp}(AA') \neq \mathrm{sp}(A'A),$$

where the prime denotes the operation of transposition.

36. Condition (54) cannot of course be fulfilled by a finite matrix. That it can be satisfied by a point A of Θ, is shown by Toeplitz's example of the matrix, A, which belongs to the infinite bilinear form $x_1y_2 + x_2y_3 + x_3y_4 + \cdots$. In fact, (54) holds for this A, since $A'A = I$ but $AA' = 0 + I$.

Accordingly, it is sufficient to prove the following theorem:

(V) *A point A of Θ cannot be in the closure of $\Theta - [\Theta]$ if it satisfies* (54).

First, if H and H_n, where $n = 1, 2, \cdots$ are bounded Hermitian matrices, then, as $n \to \infty$,

$$\text{(55)} \qquad |H_n - H| \to 0 \text{ implies that } \lim \mathrm{sp}(H_n) = \mathrm{sp}(H),$$

where "$\lim \mathrm{sp}(H_n) = \mathrm{sp}(H)$" is meant to symbolize the following situation:[4] A number, say λ, is in $\mathrm{sp}(H)$ if and only if it is possible to find in every $\mathrm{sp}(H_n)$ some number, say λ_n, in such a way that $\lambda_n \to \lambda$ as $n \to \infty$.

---

[4] The truth of the implication (55), which may be well-known, follows readily in terms of spectral resolutions from the following obvious fact: If $x_1, x_2, \cdots$ denotes a sequence of unit vectors and if μ is any real number, then the limit relation $|(H - \mu I)x_n| \to 0$, as $n \to \infty$, holds if and only if $|(H_n - \mu I)x_n| \to 0$.

In this connection, cf. P. Hartman and A. Wintner, vol. 71 (1949), pp. 865-878 of this Journal.

Next, if A is in $\Theta - [\Theta]$, and if B denotes the point A^{-1} of $\Theta - [\Theta]$, then B transforms AA' into $BAA'B^{-1}$, which is $A'A$. Since the spectrum of a bounded matrix remains invariant under transformation by a bounded matrix which has a unique bounded reciprocal matrix, it follows that

$$\text{(56)} \qquad \operatorname{sp}(AA') = \operatorname{sp}(A'A)$$

if A is in $\Theta - [\Theta]$.

37. In order to prove (V), § 36, suppose that it is false. Then there exists on Θ a point A satisfying (54) and having the property that $|A_n - A| \to 0$, as $n \to \infty$, holds for a sequence of points $A_1, A_2, \cdots$ contained in $\Theta - [\Theta]$. Hence, by the end of § 36,

$$\operatorname{sp}(A_n A'_n) = \operatorname{sp}(A'_n A_n)$$

holds for every n. On the other hand, $|A_n - A| \to 0$ implies that $|A'_n - A'| \to 0$ and that both product relations $|A_n A'_n - AA'| \to 0$, $|A'_n A_n - A'A| \to 0$ are true. Hence, if (55) is applied to both product sequences, it follows that

$$\lim \operatorname{sp}(A_n A'_n) = \operatorname{sp}(AA'), \qquad \lim \operatorname{sp}(A'_n A_n) = \operatorname{sp}(A'A).$$

But the last two formula lines imply (56) for the given A. The latter was, however, supposed to satisfy (54). This contradiction proves (V), § 36.

38. The result can be expressed as follows: Condition (56) is necessary [5] in order that a point A of Θ be in the closure of $\Theta - [\Theta]$.

It is known that, if A is any finite matrix, then AA' and $A'A$ are orthogonally equivalent, which is more than (56). One might therefore expect that, for a point A of $\Theta - [\Theta]$, the necessary condition (56) can be improved to the orthogonal equivalence of AA' and $A'A$. But this refinement of (56) proves to be false. A counter-example is supplied by the matrix, A, of the bilinear form $\epsilon_1 x_1 y_2 + \epsilon_2 x_2 y_3 + \epsilon_3 x_3 y_4 + \cdots$, if $\epsilon_1, \epsilon_2, \cdots$ is any sequence of non-vanishing numbers satisfying $\epsilon_n \to 0$ as $n \to \infty$.

In fact, the latter condition implies that A is completely continuous. This in turn implies (cf. below) that A is in the closure of $\Theta - [\Theta]$. Nevertheless, AA' and $A'A$ are not orthogonally equivalent, since they are two diagonal matrices one of which does, while the other does not, contain 0 in the diagonal, since $\epsilon_n^2 \neq 0$.

[5] That condition (56) is not sufficient as well is shown by the simple example, pointed out to us by Professor Hartman, which results if the example of Toeplitz, referred to in § 36, is bordered by a row of zeros and a column of zeros.

A class substantially more general than that of all real completely continuous matrices is defined by the following requirement for a point A of Θ: There does not exist any positive $a = a(A)$ having the property that $(\lambda I - A)^{-1}$ fails to exist (as a unique, bounded reciprocal) for every λ satisfying $|\lambda| < a$. Then $\lambda_n I - A$ is in $\Theta - [\Theta]$ for certain $\lambda_n \to 0$ as $n \to \infty$. Since this implies that $A_n = A - \lambda_n I$ satisfies $|A_n - A| \to 0$ on Θ, it follows that A is in the closure of $\Theta - [\Theta]$ (hence, in particular, (56) is satisfied). Cf. [3], p. 241.

Another sufficient condition in order that a point A of Θ be in the closure of $\Theta - [\Theta]$ consists in $AA' = A'A$. In fact, A is then normal, hence such as to have a spectral matrix, from which the existence of a sequence $A_1, A_2, \cdots$ contained in $\Theta - [\Theta]$ and satisfying $|A_n - A| \to 0$ can readily be concluded.

39. Let Γ be the metric space of *all* (real or *complex*) bounded matrices A, the distance between two points, A_1 and A_2, of Γ being defined as the least upper bound of the length of the vector $(A_1 - A_2)y$ when y varies over all unit vectors of the complex Hilbert space, and let $[\Gamma]$ denote the subset of Γ consisting of those matrices A for which there does not exist a (unique) bounded reciprocal, A^{-1}. Then it is clear from the above proofs that theorems

(α), (β) *of* § 35 *and* (V) *of* § 36 *remain true if* Θ, $[\Theta]$ *and* (54) *are replaced by* Γ, $[\Gamma]$ *and* $\mathrm{sp}(AA^*) \neq \mathrm{sp}(A^*A)$, *respectively,*

where A^* denote the complex conjugate of A'.

THE INSTITUTE FOR ADVANCED STUDY
AND
THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. H. Martin, "On infinite orthogonal matrices," *American Journal of Mathematics*, vol. 54 (1932), pp. 579-631.

[2] C. R. Putnam and A. Wintner, "The connectedness of the orthogonal group in Hilbert space," *Proceedings of the National Academy of Sciences*, vol. 37 (1951), pp. 110-112.

[3] A. Wintner, "Zur Theorie der beschränkten Bilinearformen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 228-282.

[4] ———, *Spektraltheorie der unendlichen Matrizen*, Leipzig, 1929.

ON PERTURBATIONS OF THE CONTINUOUS SPECTRUM OF THE HARMONIC OSCILLATOR.*

By PHILIP HARTMAN and AUREL WINTNER.

1. Let $f(t)$ be a real-valued continuous function on the half-line $0 \leqq t < \infty$. Consider on the latter the differential equation

$$x'' + (\lambda + f)x = 0 \tag{1}$$

with a homogeneous, linear boundary condition at $t = 0$, such as $x(0) = 0$ or, more generally,

$$x(0) \cos \alpha + x'(0) \sin \alpha = 0, \tag{2}$$

where $0 \leqq \alpha < \pi$. If (1) is of *Grenzpunkt* type, that is, if (1) and (2) determine a self-adjoint problem on the $L^2(0, \infty)$-space (for some and/or every α), let S_α denote the spectrum of this problem. Finally, let C_α denote that subset of S_α which represents the continuous spectrum.

The latter will be meant in Hilbert's sense, that is, in terms of the λ-set of non-constancy of the sum of the two continuous components (if any) of the spectral resolution analyzed into its three Lebesgue components, which are absolutely continuous, continuous but purely singular and purely discontinuous, respectively. No example of a self-adjoint problem (1)-(2) seems to be known in which the second of these three components is present.

In the case of the harmonic oscillator

$$x'' + \lambda x = 0, \tag{3}$$

that is, in the case $f(t) \equiv 0$ of (1), the explicit form of the general solution shows that

(4) (1) is of *Grenzpunkt* type

and that every non-negative λ is in C_α, while no negative λ is in the closure of S_α; so that, in particular,

$$S_\alpha' = [0, \infty) \tag{5}$$

where S_α' is the closure of S_α. In what follows, there will be delimited for an arbitrary "perturbation," $f(t)$, of (3) that "degree of smallness" (for

* Received November 16, 1949; revised October 4, 1951.

large t) under which (4) and at least the following weakened form of (5) prove to be true:

(6) $$S_a' \supset [0, \infty) \text{ in } (1).$$

It should be mentioned that the closure, S_a', of S_a, the so-called essential spectrum of (1), is always independent of a (whenever (1) holds); cf. [7], p. 251 and, for direct characterizations of this a-invariant λ-set, [3], [3 bis].

2. A sufficient "degree of smallness" turns out to be the following specification of a "small" f:

(7) $$\int_0^T |f(t)|\, dt = o(T)$$

or, more generally,

(8) $$\liminf_{T\to\infty} T^{-1} \int_0^T |f(t)|\, dt = 0.$$

In other words, (8) implies both (4) and (6). This criterion proves to be of a final nature, in the sense that (8) cannot be relaxed to

(9) $$\liminf_{T\to\infty} T^{-1} \int_0^T |f(t)|\, dt \leqq c < \infty,$$

nor even to

(10) $$\limsup_{T\to\infty} T^{-1} \int_0^T |f(t)| dt < \infty,$$

since (10) is insufficient for (6); cf. the end of Section **7**. On the other hand, it can be concluded from [4] that (10), and even (9), is sufficient for (4) alone.

3. It is natural to ask why to consider (6), instead of the sharper statement (5). The answer is that (5) can fail to be true if (8), or for that matter (7), is satisfied. In fact, it will be shown in Section **6** that not only the negation of (5) but even *the possibility of*

(11) $$S_a' = (-\infty, \infty)$$

can be realized if only (7) *is assumed.* Thus, while (6) can,

(12) $$S_a' \subset [0, \infty)$$

cannot, be concluded from (8) or (7).

The situation is changed entirely if (7) is replaced by the more drastic assumption

$$f(t) = o(1), \tag{13}$$

where $t \to \infty$. In fact, it was shown in [2] that (13) implies (5). But the proofs of the implication (13) $\to$ (5) in [2] and [5] are quite involved. A simple proof follows from the general result in [3]. It will be shown under (i) in Section **7** below that (13) $\to$ (5) can also be proved very simply by an appeal to standard theorems in the theory of operators in Hilbert space.

4. As mentioned above, both (8) $\to$ (4) and (8) $\to$ (6) are true but in the second of these conclusions, the hypothesis (8) cannot be relaxed to (10). But it turns out that (8) $\to$ (6) can be improved in a direction which relaxes the hypothesis (8) to (9), with a *fixed* $c \geqq 0$ and in such a way that the conclusion (6) becomes replaced by one depending on the numerical value of c and leading to (6) when $c = 0$.

In order to formulate this refinement of (8) $\to$ (6), it will be convenient to use the following definition: With reference to a fixed positive number c, a differential equation (1) has the property (c) if (4) is satisfied and if *every* interval of the form

$$[\lambda, \lambda + 4c + 4c^2/\lambda], \text{ where } \lambda > 0, \tag{14}$$

contains at least one point of S_α'. Then the generalization in question can be formulated as follows: *For any given* $c \geqq 0$, *property* (c) *holds whenever* (9) *is satisfied.* This will be seen at the end of Section **5**.

Except for the upper end point of the interval in (14), which is sharper in [5], the last italicized statement is a generalization of a result, proved in [5], which claims the corresponding property (c) under the assumption that

$$|f(t)| \leqq c, \tag{15}$$

rather than just (9), is satisfied.

5. The proofs will be based on the following fact: If $f(t)$ is real-valued and continuous, and if $N(T, \lambda)$ denotes the number of zeros of a (real-valued) solution $x(t) = x_\lambda(t) \not\equiv 0$ of (1) on the interval $0 \leqq t \leqq T$, then the inequality

$$|\pi N(T, \lambda) - \lambda^{\frac{1}{2}} T| \leqq 2\pi + \lambda^{-\frac{1}{2}} \int_0^T |f(t)|\, dt \tag{16}$$

holds for every $\lambda > 0$ and for every $T > 0$.

In order to see this, use will be made of the fact that, since $x(t)$ and $x'(t)$ cannot vanish simultaneously, the relation

$$(17)\qquad \theta(t) = \operatorname{arc\,tan}\{\lambda^{\frac{1}{2}}x(t)/x'(t)\}$$

and the choice of an initial determination $(\operatorname{mod}\pi)$ at $t=0$ (such as $0 \leqq \theta(0) < \pi$) define a unique *continuous* function $\theta = \theta(t)$, and that the latter has a derivative $\theta'(t)$, given by

$$(18)\qquad \theta' = \lambda^{\frac{1}{2}} + f\lambda^{\frac{1}{2}}x^2/(\lambda x^2 + x'^2).$$

This follows from (17) by virtue of (1).

If $\lambda > 0$, it follows from (18) that

$$(19)\qquad |\,\theta'(t) - \lambda^{\frac{1}{2}}\,| \leqq \lambda^{-\frac{1}{2}}\,|\,f(t)\,|.$$

On the other hand, it is seen from (17) that $\theta(t)$ becomes an integral multiple of π at exactly those t-values which are zeros of $x(t)$, while (18) shows that $\theta'(t) > 0$ whenever $x(t) = 0$. Hence it is seen from the definition of $N(T,\lambda)$ that

$$(20)\qquad |\,\pi N(T,\lambda) - \theta(T) + \theta(0)\,| \leqq 2\pi.$$

In order to obtain (16), it is sufficient to combine (20) with the inequality which results if (19) is integrated between $t=0$ and $t=T$.

Proof of (8) $\rightarrow$ (4). It follows from (8) and (16) that

$$(21)\qquad \liminf_{T\to\infty} N(T,\lambda)/T$$

has the value $\lambda^{\frac{1}{2}}/\pi$. In particular, (21) is distinct from ∞. According to criterion (**) of [4], p. 207, this is sufficient for (4).

Proof of (8) $\rightarrow$ (6). Let $0 < \lambda < \mu$. Then, by (16),

$$(22)\qquad |\,\pi[N(T,\mu) - N(T,\lambda)] - (\mu^{\frac{1}{2}} - \lambda^{\frac{1}{2}})T\,| \leqq 4\pi + 2\lambda^{-\frac{1}{2}}\int_0^T |\,f(t)\,|\,dt.$$

It follows therefore from (8) that

$$(23)\qquad \limsup_{T\to\infty}\,[N(T,\mu) - N(T,\lambda)] = \infty.$$

But Sturm's separation theorem implies that the inequality

$$N(t,\mu) - N(t,\lambda) \geqq N(T,\mu) - N(T,\lambda) - 1$$

holds for every $t > T$. Hence it is clear that the "lim sup" in (23) can be replaced by a "lim," i. e., that

$$(24) \qquad N(T, \mu) - N(T, \lambda) \to \infty \text{ as } T \to \infty.$$

Finally, it was shown in [3], p. 915, that the latter condition, (24), is sufficient in order that at least one point of S_a' be contained in the interval $[\lambda, \mu]$. Since (24) holds whenever $0 < \lambda < \mu$, the proof of (6) is complete.

Proof of (9) $\to$ (*c*). This implication (i. e., the italicized assertion of Section **4**) follows from (22) in the same way as (8) $\to$ (6) did. In fact, (9) implies (23) if $\mu^{\frac{1}{2}} - \lambda^{\frac{1}{2}} \geqq 2c/\lambda^{\frac{1}{2}}$, that is, if $\mu > \lambda + 4c + 4c^2/\lambda$.

6. The italicized statement of Section **3** concerning (11) will be proved by an example.* To this end, let $f(t)$ be a function which is continuous for $0 \leqq t < \infty$, satisfies (7) and has the constant value m on the interval

$$(25) \qquad m^4 \leqq t \leqq m^4 + m,$$

where $m = 1, 2, \cdots$. Such functions exist, since the contribution of the intervals (25) to the integral occurring in (7) is $o(T)$, simply because

$$\sum_{m=1}^{M} \int_{m^4}^{m^4+m} | f(t) | \, dt = \sum_{m=1}^{M} m^2 = o(M^4).$$

Let λ be any point of the line $-\infty < \lambda < \infty$. Since $f(t) \equiv m$ on the interval (25), a solution of (1) on (25) is $x(t) = \cos (m + \lambda)^{\frac{1}{2}} t$ whenever m is so large that $m + \lambda > 0$. Consider only such values of m. Then, by Sturm's comparison theorem, the contribution of the interval (25) to the number $N(T, \lambda)$ is between the bounds

$$\{(m + \lambda)^{\frac{1}{2}} m - 2\pi\}/\pi \text{ and } \{(m + \lambda)^{\frac{1}{2}} + 2\pi\}/\pi$$

whenever $T \geqq m^4 + m$. Hence, if $\mu > \lambda$, the contribution of such an interval (25) to the difference in (24) is

$$\geqq \{[(m + \mu)^{\frac{1}{2}} - (m + \lambda)^{\frac{1}{2}}] m - 4\pi\}/\pi.$$

Since the last $\{ \quad \}$ is $\sim \frac{1}{2}(\mu - \lambda) m^{\frac{1}{2}}$ as $m \to \infty$, it follows that (24) is satisfied. Hence, if (I) in [3], p. 915, is applied in the same way as at the end of the Proof of (8) $\to$ (6) (Section **5**), then, since $[\lambda, \mu]$ is now *any* interval, (11) follows.

---

* In a similar context, a corresponding example was communicated to us by Dr. C. R. Putnam.

7. The following remarks, (i) and (ii), do not contain new results, but simple proofs of known theorems.

(i) Suppose that $f(t)$ satisfies (13). Then the difference of the two self-adjoint operators on $L^2(0,\infty)$ which are defined by a common boundary condition (2) and by the differential operators corresponding to (1) and (3), respectively, is completely continuous. In fact, the difference of these two self-adjoint operators is represented by the operator $f(t)x(t)$ which, in view of (13) (and of the continuity of f), is completely continuous on the x-space $L^2(0,\infty)$. But (5) is true in the case (3). It follows therefore from a classical theorem ([6], p. 384; cf. [1], p. 120) that (5) holds in the case (13) of (1) also.

(ii) Besides (1), consider another differential equation of the form (1), say

$$x'' + (\lambda + g)x = 0, \tag{26}$$

where $f(t)$ and $g(t)$ are real-valued and continuous for $0 \leqq t < \infty$. It is well-known that, if

$$| g(t) - f(t) | < \text{const.}, \qquad (0 \leqq t < \infty), \tag{27}$$

then (1) must be of *Grenzpunkt* type whenever (26) is. In fact, this can be refined to an explicit asymptotic connection (as $t \to \infty$) between the general solution of (26) and that of (1), without assuming more than (27) and the negation of (4); see [8]. If this refinement is not required, a simple proof can be obtained along the lines of the argument applied under (i) above. In fact, what corresponds to the deviation $f(t)x(t)$ in (i) is now the difference

$$\{f(t) - g(t)\}x(t); \tag{28}$$

cf. (1) and (26). But (27) assures that (28) is a bounded operator on the x-space $L^2(0,\infty)$. Hence it is sufficient to apply a classical theorem ([1], p. 78), according to which the sum of a self-adjoint and of a bounded operator is always self-adjoint.

The proof in (i) was based on a general criterion concerning the perturbation of a self-adjoint operator by a completely continuous one. It may be mentioned that this criterion can be generalized to a theorem on the perturbation of the spectrum of a self-adjoint operator by the addition of an arbitrary bounded self-adjoint operator (having a norm not exceeding a given value, $c > 0$). The general theorem in question implies, as a special case, the result of [5] concerning the spectra of differential operators associated with (1) and (26), subject to the condition (27).

If (ii) is applied to $g = 0$, then, since (1), (26) reduce to (1), (3), respectively, it follows that (15) is sufficient for (4). On the other hand, (6) does not follow from (15), or even from (15) and

$$\liminf_{T\to\infty} |f(t)| = 0 \tag{29}$$

together. This is seen by choosing $f(t) = a \cos t$, where a is any non-vanishing constant. In fact, (1) then becomes Mathieu's equation, for which the positive λ-values not contained in S_a' are known to form a sequence of intervals (" instability regions ") which cluster at $\lambda = \infty$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] T. Carleman, *Sur les équations intégrales singulières á noyau réel et symétrique*, Uppsala, 1923.

[2] P. Hartman, " On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[3] ———, "A characterization of the spectra of one-dimensional wave equations," *ibid.*, pp. 915-920.

[3 bis] ——— and A. Wintner, " On the essential spectra of singular eigenvalue problems," *ibid.*, vol. 72 (1950), pp. 545-564.

[4] ——— and A. Wintner, "A criterion for the non-degeneracy of the wave equations," *ibid.*, pp. 206-213.

[5] C. R. Putnam, " The cluster spectra of bounded potentials," *ibid.*, pp. 612-620.

[6] H. Weyl, " Ueber beschränkte quadratische Formen, deren Differenz vollstetig ist," *Rendiconti del Circolo Matematico di Palermo*, vol. 27 (1909), pp. 373-392.

[7] ———, " Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[8] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

ON THE GENUS OF THE FUNDAMENTAL REGION OF SOME SUBGROUPS OF THE MODULAR GROUP.*

By Emil Grosswald.

1. The purpose of this paper is to use uniformly a general method, in order to establish the genus of the fundamental region for congruence subgroups (modulo primes) of the modular group. The results for the subgroups $\Gamma(p)$ and $\Gamma_0{}^0(p)$ are known (see [2], p. 249, and [4], p. 832, respectively); those for $\Gamma_0(p)$ are believed to be new. Let G be a group of transformations, R a simply connected fundamental region for an automorphic function admitting the group G. Let $2Q$ be the number of sides of R, C the number of cycles of its corners and P the genus of R. Then

$$P = \tfrac{1}{2}(Q - C + 1). \tag{1}$$

(see [1], p. 239, also [3, I], p. 262, (2)). Let m be the total number of independent generators and ρ the number of parabolic generators of G; let ν be the number of independent defining relations satisfied by the generators of G. Furthermore, let n be the number of cycles of R, corresponding to fixed points, μ the number of remaining cycles (corresponding to "accidental" corners). Then $n = \nu + \rho$ and the following relations hold (see [3, I], p. 170 and p. 262):

(2) $\quad Q = M;$ $\qquad$ (3) $\quad C = n + \mu = \nu + \rho + \mu.$

2. Let Γ be the group of nonhomogeneous modular transformations. Its elements are the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where the integers a, b, c, d satisfy $ad - bc = 1$ and where we consider $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$ as identical elements. Γ admits the system of generators $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

3. For every prime number p we define the subgroups $\Gamma_0(p)$ of Γ by the additional condition $c \equiv 0 \pmod{p}$. These subgroups have been studied by H. Rademacher [5], who showed that, if the square brackets stand for

* Received June 14, 1950; in revised form May 23, 1951.

the greatest integer function and the round brackets are Legendre symbols, then, for $p > 3$,

i) $\Gamma_0(p)$ is generated by $m = 2[p/12] + 3$ independent generators, satisfying $\nu = 2 + (-1/p) + (-3/p)$ defining relations;

ii) one of the generators of $\Gamma_0(p)$ is S^p, the others are of the form $V_k = \begin{pmatrix} -k_* & -1 \\ kk_* + 1 & k \end{pmatrix}$, with

$$kk_* + 1 \equiv 0 \pmod p, \qquad 1 < k, k_* < p - 1. \tag{4}$$

From (4) follows immediately the following

LEMMA. *For every p, $\Gamma_0(p)$ has exactly one parabolic generator.*

Proof. S^p is parabolic for every p. For $p = 2$ and $p = 3$, we verify directly that $V_1 = \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix}$, $V_2 = \begin{pmatrix} -1 & -1 \\ 3 & 2 \end{pmatrix}$, respectively, are not parabolic. If $p > 3$ and V_k is parabolic, it follows (see [1], p. 21) successively, by (4), that $k - k_* = \pm 2$, or $k^2 - kk_* = \pm 2k$, so that $(k \pm 1)^2 \equiv 0$ $(\mathrm{mod}\, p)$, contrary to (4), proving the lemma.

With the indicated values of m and ν, it follows from (2), (3) and the lemma, that $Q = 2[p/12] + 3$ and $n = \{2 + (-1/p) + (-3/p)\} + 1$. In the case of groups of modular transformations, $\mu = 1$ (see [3, I], p. 262 and 319) so that (1) becomes $P = [p/12] - \frac{1}{2}\{(-1/p) + (-3/p)\}$. Observing that the large bracket is a periodic function (mod 12) of p, we can write the last relation also as $P = [p/12] + r(r^2 - 25)/24$, where $r \equiv p \pmod{12}$, $|r| \leqq 5$.

4. The subgroups $\Gamma_0^0(p)$ are defined, by adding to the previous condition $c \equiv 0 \pmod p$, the new condition $b \equiv 0 \pmod p$. For $p > 3$, the values of m, ν and ρ are respectively (see [4]), $m = 2[(p+2)(p-1)/12] + 3$, $\nu = 2 + (-1/p) + (-3/p)$ and $\rho = 3$. Substituting these values in (2) and (3) we obtain from (1), $P = [(p+2)(p-1)/12] - 1 + r(r^2 - 25)/24$, where r is defined as before.

5. The principal subgroups $\Gamma(p)$ are defined, by adding to the definition of $\Gamma_0^0(p)$ the further conditions $a \equiv d \equiv 1 \pmod p$. Their structure has been studied by H. Frasch [2] and J. Nielsen [6]. H. Frasch showed that, for $p > 3$, $\Gamma(p)$ is generated as a free group by $m = p(p^2 - 1)/12 + 1$ independent generators; one of them is S^p, the others depend on three parameters. The fundamental region R', to which Frasch's generators correspond, is not simply connected; therefore, we cannot apply (1), (2) and (3) directly, but have first to transform R' into a simply connected fundamental region R. In this transformation, the number of sides does not change, so that $Q = m = p(p^2 - 1)/12 + 1$. Furthermore, $\Gamma(p)$ being a free group, it contains no elements of finite order. Consequently, the only cycles of R are those corresponding to transforms of S. As the index of $\Gamma(p)$ in Γ is $j = \frac{1}{2}p(p^2 - 1)$ and $S^h \varepsilon \Gamma(p)$ if, and only if $p \mid h$, it follows (see [2], p. 248) that $C = j/p = \frac{1}{2}(p^2 - 1)$. From (1) follows now the (well known) value

$$P = \tfrac{1}{2}\{p(p^2 - 1)/12 + 1 - \tfrac{1}{2}(p^2 - 1) + 1\} = (p^2 - 1)(p - 6)/24 + 1.$$

6. So far we have considered only the cases $p > 3$. If $p = 2$, or 3, a direct examination yields the well-known result, that the genera of $\Gamma_0(p)$, $\Gamma_0^0(p)$ and $\Gamma(p)$ are all zero.

UNIVERSITY OF SASKATCHEWAN.

BIBLIOGRAPHY.

[1] L. R. Ford, *Automorphic Functions*, New York, 1929.

[2] H. Frasch, "Die Erzeugenden der Hauptkongruenzgruppen für Primzahlstufen," *Mathematische Annalen*, vol. 108 (1933), pp. 229-252.

[3] R. Fricke-F. Klein, *Vorlesungen über die Theorie der automorphen Functionen*, vol. I (Leipzig, 1897) and vol. II (Leipzig, Berlin, 1912).

[4] E. Grosswald, "On the structure of some subgroups of the modular group," *American Journal of Mathematics*, vol. 72 (1950), pp. 809-834.

[5] H. Rademacher, "Ueber die Erzeugenden von Kongruenzuntergruppen der Modulgruppe," *Abhandlungen aus dem Seminar der Hamburgischen Universität*, vol. 7 (1929), pp. 134-138.

[6] J. Nielsen, "A study concerning the congruence subgroups of the modular group," *Det Kgl, Danske Videnskabernes Selskab Matematisk-Fysiske Meddelelsen*, vol. 25, no. 18 (1950).

CYCLOTOMY AND JACOBSTHAL SUMS.*

By Albert Leon Whiteman.

1. Introduction. Let p be an odd prime and e be a divisor of $p-1$. The Jacobsthal sum $\phi_e(n)$ is defined by

$$(1.1) \qquad \phi_e(n) = \sum_{h=1}^{p-1}\left(\frac{h}{p}\right)\left(\frac{h^e+n}{p}\right),$$

where the symbol (h/p) denotes the quadratic character of h with respect to p. In [5] Jacobsthal studied $\phi_2(n)$ and deduced an interesting application in connection with the representation of a prime p of the form $4m+1$ as the sum of two squares. Later writers [2], [6], [7], [9] obtained analogous results for other values of e. These results exhibit a connection between the Jacobsthal sum and a partition of p into quadratic summands. This connection is made precise in Theorem 1 of the present paper (§ 4). The first part of the paper (§§ 2-4) contains an account of the arithmetic properties of the sum $\phi_e(n)$ and the related sum $\psi_e(n)$ defined in § 3. There is given in § 5 a result which expresses the Jacobsthal sum in terms of a certain cyclotomic function. Finally the remainder of the paper (§§ 6-9) contains numerous applications.

2. The Jacobsthal sum. Let g be a fixed primitive root of p and write $p-1=ef$. Assume first that n in (1.1) is divisible by p. Then we see that $\phi_e(n)=0$ or $p-1$ according as e is even or odd. Next let n be prime to p, so that n is congruent to a power g^m, $m=0,1,\cdots,p-2$, of the primitive root g. From the definition it is clear that the value of $\phi_e(g^m)$ depends upon the primitive root g employed except when $m=0$. We now show that $\phi^2{}_e(g^m)=\phi^2{}_e(g^{m'})$ if $m\equiv m' \pmod e$. This result is included in the formula

$$(2.1) \qquad \phi_e(g^{et+k}) = (-1)^{t(e+1)}\phi_e(g^k),$$

where $0\leqq t\leqq f-1$, $0\leqq k\leqq e-1$. To prove (2.1) let $\bar{h}$ denote any solution of the congruence $h\bar{h}\equiv 1 \pmod p$. Then we have

$$\sum_{h=1}^{p-1}\left(\frac{h}{p}\right)\left(\frac{h^e+g^{et+k}}{p}\right)=\left(\frac{g^t}{p}\right)^{e+1}\sum_{h=1}^{p-1}\left(\frac{\bar{h}g^t}{p}\right)^{e+1}\left(\frac{1+(\bar{h}g^t)^e g^k}{p}\right)$$

$$= (-1)^{t(e+1)}\sum_{h=1}^{p-1}\left(\frac{h}{p}\right)\left(\frac{h^e+g^k}{p}\right).$$

* Received July 18, 1950.

A formula closely related to (2.1) is

$$\phi_e(nx^e) = (x/p)^{e+1}\phi_e(n) \qquad (p \nmid x). \tag{2.2}$$

This may be directly established.

Suppose now that f is odd. Then e must be even. Since e is a divisor of $p-1$, we may select x in (2.2) so that x belongs to the exponent $e \pmod p$. Then employing Euler's criterion, we get $(x/p) \equiv x^{(p-1)/2} = (x^{e/2})^f \equiv (-1)^f \equiv -1 \pmod p$. Hence in this case $\phi_e(n) = -\phi_e(n)$. We have therefore proved

$$\phi_e(n) = 0 \qquad (f \text{ odd}). \tag{2.3}$$

We next prove that

$$\phi_e(n) \equiv 0 \pmod e \qquad (f \text{ even}). \tag{2.4}$$

The congruence $h^e \equiv 1 \pmod p$ has e incongruent roots since e is a divisor of $p-1$. Let r belong to the exponent $e \pmod p$. Then for a fixed value of x not divisible by p, the e roots of the congruence $h^e \equiv x^e \pmod p$ are given by $h_i \equiv r^i x \pmod p$, $i = 0, 1, \cdots, e-1$. Now for f even we have $(r/p) \equiv r^{(p-1)/2} = (r^e)^{f/2} \equiv 1 \pmod p$. Hence $(h_i/p) = (x/p)$. We may now deduce (2.4) at once from (1.1).

Finally we establish the congruence

$$\phi_e(n) \equiv -\sum_{j=0}^{[(e-1)/2]} \binom{(p-1)/2}{(2j+1)(p-1)/2e} n^{(p-1)(e-2j-1)/2e} \pmod p \ (f \text{ even}). \tag{2.5}$$

Since f is even, $p-1$ is divisible by $2e$. Using (1.1) and Euler's criterion we expand $(h^{e+1} + nh)^{(p-1)/2}$ by the binomial theorem and interchange signs of summation; the result is

$$\phi_e(n) \equiv \sum_{v=0}^{(p-1)/2} \binom{(p-1)/2}{v} n^{(p-1)/2-v} \sum_{h=1}^{p-1} h^{(p-1)/2+ev} \pmod p.$$

In order to complete the proof we make use of the formula

$$\sum_{h=1}^{p-1} h^s \equiv \begin{cases} -1 \pmod p & (s \equiv 0 \pmod{p-1}), \\ 0 \pmod p & (s \not\equiv 0 \pmod{p-1}), \end{cases}$$

and note that $(p-1)/2 + ev$ is divisible by $p-1$ if and only if ev is an odd multiple of $(p-1)/2$.

3. The sum $\Psi_e(n)$. Related to the Jacobsthal sum $\phi_e(n)$ is the sum $\psi_e(n)$ defined as follows:

$$\psi_e(n) = \sum_{h=1}^{p-1} \left(\frac{h^e + n}{p}\right). \tag{3.1}$$

If n is divisible by p we have immediately $\psi_e(n) = p - 1$ or 0 according as e is even or odd.

For $e = 2$ we have the following formula

$$\psi_2(n) = p - 1 - (n/p) - (n/p)^2 p, \tag{3.2}$$

where the symbol (n/p) is defined as 0 when n is divisible by p. Formula (3.2) is well-known and may be proved without much difficulty. Note that $\psi_2(n)$ is equal to $p - 1$ when n is divisible by p and is equal to $-1 - (n/p)$ otherwise.

In the sequel we shall encounter the sum

$$f(a, b, c) = \sum_{x=1}^{p-1} \left(\frac{ax^2 + bx + c}{p}\right) \qquad (p \nmid a), \tag{3.3}$$

for which we have the relation

$$f(a, b, c) = -(c/p) + (-aD/p) + (a/p)\psi_2(-D) \qquad (D = b^2 - 4ac). \tag{3.4}$$

The following two formulas are analogous to (2.1) and (2.2) and are proved in much the same way.

$$\psi_e(g^{et+k}) = (-1)^{te}\psi_e(g^k) \qquad (0 \leqq t \leqq f - 1, 0 \leqq k \leqq e - 1). \tag{3.5}$$

$$\psi_e(nx^e) = (x/p)^e \psi_e(n) \qquad (p \nmid x). \tag{3.6}$$

We may also establish the following formulas:

$$\phi_e(g^k) = \begin{cases} (-1)^{k+1}\phi_e(g^{e-k}) & (e \text{ even}), \\ (-1)^{k+1}\psi_e(g^{e-k}) & (e \text{ odd}), \end{cases} \tag{3.7}$$

$$\psi_e(g^k) = (-1)^k \psi_e(g^{e-k}) \qquad (e \text{ even}), \tag{3.8}$$

for $0 \leqq k \leqq e - 1$. For example, (3.7) follows from

$$\sum_{h=1}^{p-1} \left(\frac{h}{p}\right)\left(\frac{h^e + g^k}{p}\right) = \left(\frac{g^{e-k}}{p}\right)\left(\frac{g}{p}\right)^{e+1} \sum_{h=1}^{p-1} \left(\frac{\bar{h}g}{p}\right)^{e+1} \left(\frac{(\bar{h}g)^e + g^{e-k}}{p}\right).$$

A formula of a different nature is

$$\phi_e(n) + \psi_e(n) = \psi_{2e}(n) \qquad (f \text{ even}). \tag{3.9}$$

This may be proved by summing in (1.1) and (3.1) first with respect to the quadratic residues of p and then with respect to the quadratic non-residues. Note that since f is even, $p - 1$ is divisible by $2e$.

Finally we shall prove the following congruences for $0 \leqq k \leqq e - 1$.

$$(3.10)\qquad \psi_e(g^k) \equiv \begin{cases} e \pmod{2e} & (f \text{ even, } k = 0 \text{ or } f \text{ odd, } k \neq e/2), \\ 0 \pmod{2e} & (f \text{ even, } k \neq 0 \text{ or } f \text{ odd, } k = e/2). \end{cases}$$

To establish (3.10) we write $\psi_e(g^k)$ in the form $eA - eB$, where eA is the number of times that the symbol $((h^e + g^k)/p)$ takes on the value 1 and eB is the number of times that it takes on the value -1. On the other hand $eA + eB$ is equal to $p - 1 - e$ or $p - 1$ according as the congruence $h^e + g^k \equiv 0 \pmod{p}$, $1 \leqq h \leqq p - 1$, does or does not have solutions. Since $g^{(ef)/2} = g^{(p-1)/2} \equiv -1 \pmod{p}$, we may write this congruence in the form $h^e \equiv g^{(fe)/2+k} \bmod p)$. For f even the last congruence has solutions if and only if $k = 0$; for f odd it has solutions if and only if $k = e/2$. Eliminating B by addition we get for $0 \leqq k \leqq e - 1$

$$(3.11)\qquad \psi_e(g^k) = \begin{cases} -p + 1 + e + 2eA & (f \text{ even, } k = 0 \text{ or } f \text{ odd, } k = e/2), \\ -p + 1 + 2eA & (f \text{ even, } k \neq 0 \text{ or } f \text{ odd, } k \neq e/2), \end{cases}$$

from which (3.10) may easily be derived.

4. The main results. We first prove

$$(4.1)\qquad \sum_{k=0}^{e-1} \phi_e(g^k) = -e \qquad (e \text{ odd}).$$

To prove this we make use of (2.1) with e odd. Thus we get

$$\frac{p-1}{e} \sum_{k=0}^{e-1} \phi_e(g^k) = \sum_{t=0}^{f-1} \sum_{k=0}^{e-1} \phi_e(g^{et+k}) = \sum_{m=0}^{p-2} \phi_e(g^m)$$
$$= -(p-1) + \sum_{a=0}^{p-1} \phi_e(a) = -(p-1).$$

Equation (4.1) follows immediately.

We now state our principal result.

THEOREM 1. *Let s be a fixed integer such that $0 \leqq s \leqq e - 1$. If e is odd then*

$$(4.2)\qquad \sum_{k=0}^{e-1} \phi_e(g^k)\phi_e(g^{k+s}) = \begin{cases} e^2 p - e(p-1) & (s = 0), \\ -e(p-1) & (s \neq 0). \end{cases}$$

If e is even and f is even then

$$(4.3)\qquad \sum_{k=0}^{e-1} \phi_e(g^k)\phi_e(g^{k+s}) = \begin{cases} e^2 p & (s = 0), \\ 0 & (s \neq 0). \end{cases}$$

To prove this theorem let S denote the left member of (4.2) or (4.3).

It should be noted that when f is odd each term of S is equal to zero in view of (2.3). We get by (1.1) and (2.1)

$$\begin{aligned}
\frac{p-1}{e} S &= \sum_{t=0}^{f-1} \sum_{k=0}^{e-1} \phi_e(g^{et+k})\phi_e(g^{et+k+s}) \\
&= \sum_{m=0}^{p-2} \phi_e(g^m)\phi_e(g^{m+s}) \\
&= \sum_{x,y=1}^{p-1} \left(\frac{x}{p}\right)\left(\frac{y}{p}\right) \sum_{m=0}^{p-2} \left(\frac{g^s g^{2m} + (x^e g^s + y^e) g^m + x^e y^e}{p}\right) \\
&= \sum_{x,y=1}^{p-1} \left(\frac{x}{p}\right)\left(\frac{y}{p}\right) f(g^s, x^e g^s + y^e, x^e y^e),
\end{aligned}$$

where $f(a, b, c)$ is defined in (3.3). Making use of (3.2) and (3.4) this becomes

$$\frac{p-1}{e} S = -\sum_{x,y=1}^{p-1} \left(\frac{xy}{p}\right)^{e+1} - \left(\frac{g^s}{p}\right) p \sum_{x=1}^{p-1} \left(\frac{x}{p}\right) \sum_{y=1}^{p-1} \left(\frac{y}{p}\right)\left(\frac{y^e - x^e g^s}{p}\right)^2.$$

If $s \neq 0$ the right member of the last equation reduces to $-\sum_{x,y=1}^{p-1} (xy/p)^{e+1}$, which is equal to 0 if e is even and is equal to $-(p-1)^2$ if e is odd. If $s = 0$ we note that for a fixed value of x the value of the symbol $((y^e - x^e)/p)^2$ is equal to 0 if $y^e \equiv x^e \pmod p$ and is equal to 1 otherwise. In the proof of formula (2.4) we showed that when f is even the e roots of the congruence $y^e \equiv x^e \pmod p$ are given by $y_i \equiv r^i x \pmod p$, $i = 0, 1, \cdots, e-1$, where r belongs to the exponent $e \pmod p$ and $(r/p) = 1$. Hence

$$\begin{aligned}
\sum_{x=1}^{p-1} \left(\frac{x}{p}\right) \sum_{y=1}^{p-1} \left(\frac{y}{p}\right)\left(\frac{y^e - x^e}{p}\right)^2 &= -\sum_{x=1}^{p-1} \left(\frac{x}{p}\right) \sum_{i=0}^{e-1} \left(\frac{r^i x}{p}\right) \\
&= -\sum_{x=1}^{p-1} \left(\frac{x}{p}\right)^2 \sum_{i=0}^{e-1} \left(\frac{r}{p}\right)^i = -e(p-1).
\end{aligned}$$

Combining our results we obtain Theorem 1.

The method of this section may be used to derive corresponding results for the sum $\psi_e(n)$. Thus we get for e even or odd

$$\sum_{k=0}^{e-1} (-1)^k \psi_e(g^k) = -e, \tag{4.4}$$

which is analogous to (4.1). We may also prove

THEOREM 2. *Let s be a fixed integer such that $0 \leqq s \leqq e-1$. If e is even then*

$$(4.5)\qquad \sum_{k=0}^{e-1}\psi_e(g^k)\psi_e(g^{k+s})=\begin{cases} e^2p-2e(p-1) & (s=0),\\ -2e(p-1) & (s \text{ even}, s\neq 0),\\ 0 & (s \text{ odd}).\end{cases}$$

If e is odd then

$$(4.6)\qquad \sum_{k=0}^{e-1}\psi_e(g^k)\psi_e(g^{k+s})=\begin{cases} e^2p-e(p-1) & (s=0),\\ -e(p-1) & (s \text{ even}, s\neq 0),\\ e(p-1) & (s \text{ odd}).\end{cases}$$

5. Cyclotomy. The cyclotomic number (h,k) is the number of values of y, $1\leqq y\leqq p-2$, for which

$$(5.1)\qquad y\equiv g^{es+h},\qquad 1+y\equiv g^{et+k}\qquad (\operatorname{mod} p),$$

where the values of s and t are each selected from the integers $0,1,\cdots,f-1$.

Noting that $g^{ef}\equiv 1 \pmod p$ we may easily infer that

$$(5.2)\qquad (h,k)=(h+ae,k+be),$$

for any integers a and b. Furthermore it is not difficult to prove that

$$(5.3)\qquad (h,k)=\begin{cases}(e-h,k-h) & (f \text{ even or odd}),\\ (k,h) & (f \text{ even}),\\ (k+e/2,h+e/2) & (f \text{ odd}),\end{cases}$$

and

$$(5.4)\qquad \sum_{h=0}^{e-1}(h,k)=\begin{cases} f-1 & (k=0),\\ f & (1\leqq k\leqq e-1).\end{cases}$$

For a proof of (5.3) and (5.4) see, for example, Bachmann [1; Chapter 15].

We now consider the sum

$$(5.5)\qquad B(v,n)=\sum_{h=0}^{e-1}(h,v-nh).$$

This sum is equal to the number of values of y, $1\leqq y\leqq p-1$, for which $y^n(1+y)$ is congruent to a number of the form $x^eg^v(p\nmid x)$ with respect to the modulus p. Hence the number of solutions of the congruence

$$(5.6)\qquad y^{n+1}+y^n\equiv x^eg^v \pmod p \qquad (v \text{ fixed};\ 0\leqq x,y\leqq p-1),$$

is equal to $2+eB(v,n)$. For a fixed integer a let $F_n(a)$ denote the number of values of y, $1\leqq y\leqq p-1$, for which $y^{n+1}+y^n\equiv a \pmod p$. In terms of the functions $F_n(a)$ the number of solutions of the congruence (5.6) is equal to $\sum_{x=0}^{p-1}F_n(x^eg^v)$. Combining our results we conclude that

$$\sum_{x=0}^{p-1} F_n(x^e g^v) = 2 + eB(v, n). \tag{5.7}$$

The case $n = 1$ is an important special case. We may easily show that

$$F_1(x^e g^v) = 1 + \left(\frac{1 + 4g^v x^e}{p}\right),$$

so that (5.7) becomes

$$\sum_{x=0}^{p-1} \left(\frac{1 + 4g^v x^e}{p}\right) = -p + 2 + eB(v, 1).$$

The last equation leads at once to

$$eB(v, 1) = \begin{cases} p - 1 + \phi_e(4g^v) & (e \text{ odd}), \\ p - 1 + \psi_e(4g^v) & (e \text{ even}), \end{cases} \tag{5.8}$$

where $\phi_e(n)$ and $\psi_e(n)$ are defined by (2.1) and (3.1), respectively.

6. The case $e = 2$. As a first application let $e = 2$ and f be even, so that $p \equiv 1 \pmod 4$. In this case the second member of (4.3) reduces to the identity $p = a^2 + b^2$, where

$$a = \phi_2(1)/2, \qquad b = \phi_2(g)/2. \tag{6.1}$$

Next using (3.2), (3.9) with $e = 2$ and (3.10) with $e = 4$ we get

$$a \equiv -1 \pmod 4. \tag{6.2}$$

Formulas (6.1) and (6.2) constitute the theorem of Jacobsthal [5].

Applying (2.5) we obtain at once

$$2a \equiv -\binom{f}{f/2} \pmod p. \tag{6.3}$$

Formula (6.3) is a theorem of Gauss [3].

Again using (3.2) and (3.9) with $e = 2$ we get

$$a = (\psi_4(1) + 2)/2, \qquad b = \psi_4(g)/2. \tag{6.4}$$

Formula (6.4) is a theorem of Chowla [2; Theorem 1]. It may be remarked that (6.4) may also be deduced directly from (3.8), (4.4) and (4.5).

7. The case $e = 3$. If $p = 3f + 1$ the diophantine equation $4p = x^2 + 3y^2$ has three solutions in positive integers x and y. By (4.1) and (4.2) we get the equations

$$(7.1)\qquad \phi_3(1)+\phi_3(g)+\phi_3(g^2)=-3,\quad \phi_3{}^2(1)+\phi_3{}^2(g)+\phi_3{}^2(g^2)=6p+3,$$

which may readily be transformed into the three identities

$$(7.2)\qquad 4p=(1+\phi_3(g^i))^2+3\left(\frac{\phi_3(g^{i+1})-\phi_3(g^{i+2})}{3}\right)^2 \qquad (i=0,1,2).$$

We now prove that the three solutions of $4p=x^2+3y^2$ given by (7.2) are distinct. For this purpose it suffices to show that

$$1+\phi_3(g^i)\neq\pm(1+\phi_3(g^j)) \qquad (i\neq j;\, i,j=0,1,2).$$

If $1+\phi_3(g^i)=1+\phi_3(g^j)$ then equations (7.1) lead to the absurd conclusion that $(1+\phi_3(g^i))^2=p$. If $1+\phi_3(g^i)=-(1+\phi_3(g^j))$ then the first equation in (7.1) implies that $1+\phi_3(g^k)=0$ for $k\neq i$, $k\neq j$. Hence by (7.2) we deduce $3\,|\,4p$, which is impossible.

The diophantine equation $4p=c^2+3d^2$ has a unique solution with $c\equiv 1 \pmod 3$ and $d\equiv 0 \pmod 3$. This solution is given by

$$(7.3)\qquad c=1+\phi_3(4),\qquad d=(\phi_3(4g)-\phi_3(4g^2))/3.$$

To prove (7.3) we note that the congruence $c\equiv 1 \pmod 3$ follows at once from (2.4). Next using (5.2), (5.3), (5.5) and (5.8) we get

$$\phi_3(4g)=-p+1+9(0,1),\qquad \phi_3(4g^2)=-p+1+9(0,2),$$

so that $d\equiv 0 \pmod 3$. Formula (7.3) is the theorem of von Schrutka [7].

Applying (2.5) we obtain

$$(7.4)\qquad c\equiv -2^{2f}\binom{3f/2}{f/2}\equiv-\binom{2f}{f} \qquad (\mathrm{mod}\ p).$$

Formula (7.4) is due to Jacobi [4].

By (3.5), (3.7) and (3.9) with $e=3$ we get

$$2\phi_3(1)=\psi_6(1),\qquad \phi_3(g)-\phi_3(g^2)=\psi_6(g),$$

so that for $i=0$, (7.2) reduces to the identity $p=s^2+3t^2$, where

$$(7.5)\qquad s=(\psi_6(1)+2)/4,\qquad t=\psi_6(g)/6.$$

and the sign of s is determined by means of the congruence

$$(7.6)\qquad s\equiv -1 \qquad (\mathrm{mod}\ 3),$$

because of (3.10) with $e=6$. We have also by (2.5)

$$(7.7)\qquad 2s\equiv-\binom{3f/2}{f/2} \qquad (\mathrm{mod}\ p).$$

Formulas (7.5) and (7.6) comprise a theorem due to Chowla [2; Theorem 2]. We remark that (7.5) may also be deduced directly from (3.8), (4.4) and (4.5).

8. The case $e = 4$. In this section we consider the case $e = 4$ and f is even, so that $p \equiv 1 \pmod 8$. By (3.7) and (3.8) we get

$$\phi_4(g) = \phi_4(g^3),\ \phi_4(g^2) = 0; \qquad \phi_4^2(1) + \phi_4^2(g) + \phi_4^2(g^2) + \phi_4^2(g^3) = 16p.$$

Hence we have the identity $p = x^2 + 2y^2$, where

$$(8.1) \qquad x = \phi_4(1)/4, \qquad y = \phi_4(g)/4.$$

For another proof of (8.1) see a recent paper [9; Theorem 2].

The sign of x in (8.1) is determined by means of the congruence

$$(8.2) \qquad x \equiv (-1)^{f/2+1} \pmod 4.$$

To prove (8.2) we return to (3.11) and obtain the equation

$$\psi_4(1) = -p + 5 + 8A,$$

where $4A$ denotes the number of times that the symbol $((h^4+1)/p)$, $h = 1, 2, \cdots, p-1$, takes on the value 1. We next prove that A is an odd integer. Put $h \equiv g^k \pmod p$, $k = 0, 1, \cdots, p-2$. For $k = f/2, 3f/2, 5f/2, 7f/2$ we get $((g^{4k}+1)/p) = 0$. For $k = 0, f, 2f, 3f$ we get $((g^{4k}+1)/p) = (2/p) = 1$. For $k = 1, 2, \cdots, f/2-1, f/2+1, \cdots, f-1$ group the k's in pairs so that $f/2-j, f/2+j, j = 1, 2, \cdots, f/2-1$ form a pair. Since

$$(8.3) \qquad (1 + g^{4(f/2-j)})(1 + g^{4(f/2+j)}) \equiv g^{4(f/2-j)}(1 + g^{4(f/2+j)})^2 \pmod p,$$

it follows that the two factors in the left member of (8.3) have the same quadratic character. Therefore the number of values of k, $k = 0, 1, \cdots, f-1$, for which $((g^{4k}+1)/p)$ is equal to 1 is $2q+1$ for some integer q. Furthermore $g^{4(f/2\pm j)} \equiv g^{4(f/2+af\pm j)} \pmod p$ for any integer a. Hence $A = 2q+1$. This, in turn, leads to the congruence $\psi_4(1) \equiv (-1)^{f/2+1}\,4 \pmod{16}$. Using (3.9) with $e = 4$ and (3.10) with $e = 8$ we may now deduce (8.2).

Applying (2.5) we get easily

$$(8.4) \qquad 2x \equiv -\binom{2f}{f/2} \pmod p.$$

Formula (8.4) is a theorem of Stern [8].

9. The case $e = 5$. For a prime $p = 5f + 1$, f is even. By (4.1) and (4.2) we have the identities

$$(9.1)\qquad \begin{gathered} \sum_{k=0}^{4} \phi_5(g^k) = -5, \qquad \sum_{k=0}^{4} \phi_5{}^2(g^k) = 20p + 5, \\ \sum_{k=0}^{4} \phi_5(g^k)\phi_5(g^{k+1}) = \sum_{k=0}^{4} \phi_5(g^k)\phi_5(g^{k+2}) = -5(p-1). \end{gathered}$$

After some manipulation we may transform the identities in (9.1) into the identities

$$(9.2)\qquad \begin{gathered} 16p = A_k{}^2 + 5B_k{}^2 + 10C_k{}^2 + 10D_k{}^2 \qquad (k = 0, 1, 2, 3, 4), \\ A_kB_k = C_k{}^2 - 4C_kD_k - D_k{}^2, \end{gathered}$$

where

$$(9.3)\qquad \begin{gathered} A_k = 1 + \phi_5(g^k),\ 5B_k = \phi_5(g^{k+1}) - \phi_5(g^{k+2}) - \phi_5(g^{k+3}) + \phi_5({}^{(k+4)}), \\ 5C_k = \phi_5(g^{k+2}) - \phi_5(g^{k+3}),\ 5D_k = \phi_5(g^{k+1}) - \phi_5(g^{k+4}). \end{gathered}$$

Another pair of identities is

$$(9.4)\qquad 16p = x^2 + 50u^2 + 50v^2 + 125w^2, \qquad xw = v^2 - 4uv - u^2,$$

where

$$(9.5)\qquad \begin{aligned} x &= 1 + \phi_5(4), \\ 25u &= \phi_5(4g) + 2\phi_5(4g^2) - 2\phi_5(4g^3) - \phi_5(4g^4), \\ 25v &= 2\phi_5(4g) - \phi_5(4g^2) + \phi_5(4g^3) - 2\phi_5(4g^4), \\ 25w &= \phi_5(4g) - \phi_5(4g^2) - \phi_5(4g^3) + \phi_5(4g^4), \end{aligned}$$

and the sign of x is determined by means of the congruence $x \equiv 1 \pmod 5$ in view of (2.4).

The formula for x is due to Emma Lehmer [6]. To prove (9.4) select k in (9.3) so that $g^k \equiv 4 \pmod p$. Making use of the formulas in § 5 we may express u, v and w defined in (9.5) in terms of the cyclotomic numbers (h, k). Dropping subscripts in (9.3) we arrive at the equations

$$\begin{aligned} 25u &= 10C + 5D = 25[(0,2) - (0,3)], \\ 25v &= -5C + 10D = 25[(0,1) - (0,4)], \\ 25w &= 5B = 25[(1,3) - (1,2)]. \end{aligned}$$

Hence $A = x$, $B = 5w$, $C = 2u - v$, $D = u + 2v$ and (9.2) reduces to (9.4).

UNIVERSITY OF SOUTHERN CALIFORNIA.

REFERENCES.

[1] P. Bachmann, *Die Lehre von der Kreistheilung*, Leipzig-Berlin, 1921.

[2] S. Chowla, " The last entry in Gauss's diary," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 244-246.

[3] K. F. Gauss, " Theorie residuorum biquadraticorum," *Werke*, vol. 2, p. 90.

[4] K. Jacobi, " De residuis cubicis commentatio numersoa," *Journal für die reine und angewandte Mathematik*, vol. 2 (1827), pp. 66-69.

[5] E. Jacobsthal, " Über die Darstellung der Primzahlen der Form $4n+1$ als Summe zweier Quadrate," *Journal für die reine und angewandte Mathematik*, vol. 132 (1907), pp. 238-245.

[6] E. Lehmer, " On the quintic character of 2," *Bulletin of the American Mathematical Society*, vol. 55 (1949), Abstract No. 72, p. 62.

[7] L. von Schrutka, " Eine Beweis für die Zerlegbarkeit der Primzahlen von der Form $6n+1$ in ein einfaches und ein dreifaches Quadrat," *Journal für die reine und angewandte Mathematik*, vol 140 (1911), pp. 252-265.

[8] M. Stern, " Eine Bemerkung zur Zahlentheorie," *Journal für die reine und angewandte Mathematik*, vol. 32 (1846), pp. 89-90.

[9] A. L. Whiteman, "Theorems analogous to Jacobsthal's theorem," *Duke Mathematical Journal*, vol. 16 (1949), pp. 619-626.

INFINITELY NEAR POINTS ON ALGEBRAIC SURFACES.*

By Gino Turrin.[1]

1. Infinitely near points on plane algebraic curves were introduced by Max Noether [1] in the process of reduction of singularities. Noether's classical theorem is: a sequence of singular infinitely near points on a plane algebraic curve is necessarily finite. In the case of space curves and surfaces the first definitions and results were established by C. Segre [2] using quadratic transformations of the ambient space. Noether's result, though it remains valid for space curves, is no longer true for surfaces because of the appearance of singular curves.

As a general rule, and even in the simplest case of plane algebraic curves, the analysis of infinitely near points is intricate and all proofs have to be carried out by handling a great number of details.[2] In the case of algebraic surfaces B. Levi [3] proved the following theorem: *if $P, P_1, P_2, \cdots$ is an infinite sequence of infinitely near points on an algebraic surface and all of the same multiplicity $\nu > 1$, then for any given p there exists $q > p$ such that the point P_q lies on the transform of a ν-fold curve passing through the point P_{q-1} immediately preceding P_q.* It is the purpose of this note to give a purely algebraic proof of Levi's theorem for arbitrary ground fields of characteristic zero. Our proof makes use of related results proved in the fundamental paper of Zariski [7] and it goes further by showing the existence of an index p such that for *any* $q > p$ the point P_q lies on a ν-fold curve.

2. A trivial case of an infinite sequence of infinitely near points which all have the same multiplicity $\nu > 1$ is one in which there exists an index p such that the point P_p lies on a ν-fold curve Γ_p and all the points sucessive to P_p lie on the corresponding transforms of the curve Γ_p. A sequence of this sort will be called a *trivial sequence.*

We shall need the following remark: *Let P be a point of a surface F and let T be a quadratic transformation of center P which sends F into a surface F_1 and the fundamental point P into the curve $\Delta = T[P]$ of F.*

* Received October 5, 1950.

[1] Fellow of the Rockefeller Foundation.

[2] See for example [5] or [6], chap. 1.

If Γ_1 *is an irreducible curve of* F_1 *not contained as component in* Δ *and such that* $\Gamma_1 \cap \Delta \neq 0$, *then* Γ_1 *is the transform* $T[\Gamma]$ *of an irreducible curve* Γ *of* F *passing through* P *and* Γ_1 *has for* F_1 *the same multiplicity as* Γ *has for* F. This is immediate: since P is the only fundamental point of the transformation T and since Γ_1 is not a component of $\Delta = T[P]$ it follows that Γ_1 is the transform $T[\Gamma]$ of an irreducible curve Γ of the surface F. Since under the inverse transformation T^{-1} to any point of Δ corresponds the point P and since $\Gamma_1 \cap \Delta \neq 0$ we conclude that Γ contains the point P. Finally, the respective multiplicities of Γ and Γ_1 are the same since the quadratic transformation T is regular at a general point of the curve Γ.

3. Our result is as follows:

THEOREM. *Let* $P, P_1, P_2, \cdots$ *be an infinite sequence of infinitely near points all of the same multiplicity* $\nu > 1$ *on an irreducible algebraic surface* F *defined over an arbitrary ground field of characteristic zero. Then there exists an index* p *such that for any* $q > p$ *the point* P_q *lies on at least one irreducible curve* Γ_q *which is* ν*-fold for the surface* F_q *containing* P_q. *The curve* Γ_q *either* a) *coincides with the transform* $T_{q-1}[P_{q-1}]$ *of the point* P_{q-1} *under the quadratic transformation* $T_{q-1}\colon F_{q-1} \rightarrow F_q$, *or* b) *is the transform* $T_{q-1}[\Gamma_{q-1}]$ *of some curve* Γ_{q-1} *passing through* P_{q-1} *and* ν*-fold for the surface* F_{q-1}. *Case* b) *must occur infinitely many times.*

Proof. A given sequence $P, P_1, P_2, \cdots$ of points infinitely near to P on the surface F determines at least one valuation v of the field of algebraic functions of which the birationally equivalent surfaces $F, F_1, F_2, \cdots$ are distinct projective models. If the given infinite sequence $P, P_1, P_2, \cdots$ is a trivial sequence the Theorem is obviously true. Therefore we assume that *the sequence* $P, P_1, P_2, \cdots$ *is not trivial and all the points have the same multiplicity* $\nu > 1$. Zariski [7] has proved that under these conditions the valuation v must be 0-dimensional non-discrete of rank one and that moreover, after a finite number of quadratic transformations, the element ω_i which defines the surface F_i in the quotient ring $\mathfrak{J}_i$ of the point P_i with respect to the 3-dimensional ambient space of F_i can be written in the following form: (we may assume that $P_i = P$ and $F_i = F$ are the initial point and surface respectively)

$$\omega = \epsilon_0 z^\nu + \sum_{j=1}^{\nu} \epsilon_j z^{\nu-j} x^{m_{0j}} y^{n_{0j}} \tag{1}$$

where the ϵ_j, $0 \leqq j \leqq \nu$, are either zero or units of the quotient ring $\mathfrak{J}$ (but $\epsilon_0 \neq 0$) and where z, x, y is a set of uniformizing parameters of the point P

at $\mathfrak{I}$. Since the point $P = (0, 0, 0)$ is of multiplicity ν for the surface F the exponents which appear in (1) must satisfy the inequalities $m_{0j} + n_{0j} \geqq j$, for all $j \neq 0$ such that $\epsilon_j \neq 0$. Furthermore, *loc. cit.* Lemma 11. 4, the leading form of the element ω is the ν-th power of a linear form. Taking into account this fact and the expression (1) we shall obtain the stronger inequalities

$$m_{0j} + n_{0j} > j \tag{2}$$

for all $j \neq 0$ such that $\epsilon_j \neq 0$. For let $\mathfrak{L} = (a_1\bar{z} + a_2\bar{y} + a_3\bar{x})^\nu$ be the leading form of ω where a_1, a_2, a_3 are elements of the residue field $\mathfrak{K}$ of P, $\bar{z}, \bar{y}, \bar{x}$ are algebraically independent transcendentals over $\mathfrak{K}$, and where $a_1 \neq 0$ since $\epsilon_0 \neq 0$. If, say, $a_2 \neq 0$ then necessarily $a_3 = 0$ since otherwise the expression of $\mathfrak{L}$ as a sum of monomials would contain $(\nu + 1)(\nu + 2)/2$ $(> \nu + 1)$ terms contradicting the expression (1) for ω. Thus we may assume $a_1, a_2 \neq 0$ and $a_3 = 0$. Now the expression for $\mathfrak{L}$ as a sum of monomials has exactly $\nu + 1$ terms. This implies in (1) that all ϵ_j are different from zero and also that the total degree in the uniformizing parameters of each monomial of ω is exactly ν. Since $\bar{x}$ does not occur in $\mathfrak{L}$, the expression (1) for ω takes the form $\omega = \epsilon_0 z^\nu + \epsilon_1 z^{\nu-1} y + \cdots + \epsilon_\nu y^\nu$. Let α_1 and α_2 be elements of $\mathfrak{I}$ whose residues coincide with a_1 and a_2 respectively. Then the leading form of the element $\omega' = (\alpha_1 z + \alpha_2 y)^\nu$ coincides with $\mathfrak{L}$ and since the quotient ring $\mathfrak{I}$ of the point P with respect to the ambient 3-dimensional space is a regular local ring we conclude that $\omega = \omega'$. This contradicts the fact that F is an irreducible surface. Hence necessarily also $a_2 = 0$, i. e., $\mathfrak{L} = (a_1\bar{z})^\nu$ which, by definition of the leading form, shows the validity of the inequalities (2) for the expression (1).

We point out that the expression (1) together with conditions (2) is permanent in the sense that all the elements ω_i which succeed ω when the successive quadratic transformations are performed have expressions similar to (1) in which the corresponding conditions analogous to (2) hold. In order to write the local equations of the quadratic transformation T which sends the surface F defined in $\mathfrak{I}$ by ω into the surface F_1 defined in $\mathfrak{I}_1$ by ω_1, and thus to obtain for ω_1 the expression similar to (1) in terms of the uniformizing parameters z_1, y_1, x_1 of the point P_1, we must distinguish three cases according to the ratio of the values assigned to y and x by the fixed valuation v:

Case 1. $v(x) < v(y)$. The local equations of T are:

$$z_1 = z/x, \qquad y_1 = y/x, \qquad x_1 = x; \tag{4.1}$$

consequently

$$\omega_1 = \omega/x^\nu = \epsilon_0 z_1{}^\nu + \sum_{j=1}^{\nu} \epsilon_j z_1{}^{\nu-j} x_1{}^{m_{1j}} y_1{}^{n_{1j}} \tag{1.1}$$

where

$$m_{1j} = m_{0j} + n_{0j} - j, \qquad n_{1j} = n_{0j} \tag{3.1}$$

and where the conditions similar to (2) are

$$m_{1j} + n_{1j} > j. \tag{2.1}$$

Case 2. $v(x) = v(y)$. Since the valuation v has center in all the quotient rings $\mathfrak{J}_i$, in particular in $\mathfrak{J}_1$, the element y/x is a unit in $\mathfrak{J}_1$ and we have:

$$z_1 = z/x, \qquad y_1 = f(y/x), \qquad x_1 = x \tag{4.2}$$

where $f(y/x)$ is a suitable polynomial in y/x with coefficients in $\mathfrak{J}$.[3]

$$\omega_1 = \omega/x^\nu = \epsilon_0 z_1{}^\nu + \sum_{j=1}^{\nu} \bar{\epsilon}_j z_1{}^{\nu-j} x_1{}^{m_{1j}} \tag{1.2}$$

where $\bar{\epsilon}_j = \epsilon_j \cdot (y/x)^{n_{0j}}$ is clearly a unit in $\mathfrak{J}_1$ (if $\neq 0$) and where

$$m_{1j} = m_{0j} + n_{0j} - j, \qquad (n_{1j} = 0) \tag{3.2}$$

and

$$m_{1j} > j. \tag{2.2}$$

Case 3. $v(x) > v(y)$. In this case we have:

$$z_1 = z/y, \qquad y_1 = y, \qquad x_1 = x/y, \tag{4.3}$$

$$\omega_1 = \omega/y^\nu = \epsilon_0 z_1{}^\nu + \sum_{j=1}^{\nu} \epsilon_j z_1{}^{\nu-j} x_1{}^{m_{1j}} y_1{}^{n_{1j}}, \tag{1.3}$$

$$m_{1j} = m_{0j}, \qquad n_{1j} = n_{0j} + m_{0j} - j, \tag{3.3}$$

and as usual

$$m_{1j} + n_{1j} > j. \tag{2.3}$$

It follows from (1), (2) and (4. s) ($s = 1, 2, 3$) that the curve $\Delta = T[P]$ which corresponds on the surface F_1 to the fundamental point $P = (0, 0, 0)$ is irreducible and is given in the quotient ring $\mathfrak{J}_1$ of P_1 by the ideal (z_1, x_1) in cases 1 and 2, and by the ideal (z_1, y_1) in case 3. We shall also make use of the fact that in case 3 the curve given in $\mathfrak{J}_1$ by the

[3] If the ground field is algebraically closed $f(y/x) = y/x - c$, where c is the v-residue of y/x.

ideal (z_1, x_1) is merely the transform of the curve which passes through the point $P = (0,0,0)$ and is given in $\mathfrak{J}$ by the ideal (z, x).

With reference to the first part of the Theorem we observe first of all that *given any index p there exists $q > p$ such that the point P_q is not isolated.*[4] In fact, our hypothesis that $P, P_1, P_2, \cdots$ is a non-trivial sequence of points all of the same multiplicity $\nu > 1$ implies that the local uniformization of the 0-dimensional valuation v ([7] Theorem 5) cannot be performed by using quadratic transformations only. Since the process of uniformization of v is carried out by a finite number of quadratic and monoidal transformations and since monoidal transformations come into play only when some intermediate point P_i is not isolated, our assertion follows. Therefore we may asume that in the surface F given by (1) there is a ν-fold curve passing through the point $P = (0,0,0)$. Moreover, *we may assume that the ν-fold curve is given in the quotient ring $\mathfrak{J}$ of P by the ideal (z, x).* For if no point after P is isolated the first part of the Theorem is already proved. Otherwise since we have just shown that in our given infinite sequence of points there exists an infinite number of points which are not isolated, we may suppose for a moment that the point P is isolated and that P_1 is not isolated. By the remark of § **2** the ν-fold curve through P_1 must coincide with the transform $\Delta = T[P]$ of the point P. Now, we have seen above that Δ is given in the quotient ring $\mathfrak{J}_1$ of P_1 either by the ideal (z_1, x_1) (cases 1 and 2) or by the ideal (z_1, y_1) (case 3). Hence, by setting again $P_1 = P$ and interchanging, if necessary, the notation for the two uniformizing parameters x and y, we prove the assertion. Therefore we assume that in the surface F given by (1) the curve defined in the quotient ring $\mathfrak{J}$ by the ideal (z, x) is ν-fold for F. This is equivalent to saying that in (1) the following additional conditions to (2) hold:

$$m_{0j} \geqq j \tag{2'}$$

for all $j \neq 0$ such that $\epsilon_j \neq 0$. Once all the above assumptions are made the first part of the Theorem would be proved when it is shown that *no matter which case* (1, 2 or 3) *occurs, then in the corresponding expression for ω_1 we necessarily have $m_{1j} \geqq j$ for all $j \neq 0$ such that $\epsilon_j \neq 0$.* This is given by (2.2) in case 2 and is an obvious consequence of (2′) and the first part of (3.3) in case 3. To prove the assertion for case 1 we need the following general remark applicable to the expression (1): *for no $j \neq 0$ such that $\epsilon_j \neq 0$ can it happen that*

$$\text{(A)} \quad m_{0j} < j \qquad\qquad \text{(B)} \quad n_{0j} < j$$

[4] A ν-fold point P is isolated if no curve of the surface passing through P is ν-fold.

simultaneously. For whatever case 1, 2 or 3 occurs as the next step, it is clear from (A), (B) and (3. s) ($s = 1, 2, 3$) that

$$m_{0j} + n_{0j} > m_{1j} + n_{1j}$$

and that again, using once more (A), (B) and (3. s), we have simultaneously

$$(\mathrm{A}_1) \quad m_{1j} < j \qquad\qquad (\mathrm{B}_1) \quad n_{1j} < j.$$

Thus we should get an infinite strictly decreasing sequence of inequalities

$$m_{0j} + n_{0j} > m_{1j} + n_{1j} > m_{2j} + n_{2j} > \cdots$$

contradicting the generally valid inequalities similar to (2). Now, if in (3. 1) we had, for some $j \neq 0$ for which $\epsilon_j \neq 0$, $m_{1j} < j$ it would follow from (2′) and (3. 1) also that $n_{1j} < j$. This we have seen to be impossible and hence also in case 1 we must necessarily have $m_{1j} \geqq j$ for all $j \neq 0$ such that $\epsilon_j \neq 0$. This completes the proof of the first part of the Theorem. The second part of the Theorem is evident from the remark of § **2** and together with the first part represents our improvement of Levi's result in the sense that there exists a certain point in the sequence such that *all* its successors are non-isolated and the ν-fold curves which appear through the points are of a certain well determined type.

To prove the last part of the Theorem we shall now show that, *given an index p there exists $q > p$ such that the point P_q lies on the transform of a ν-fold curve passing through P_{q-1}.* Returning to the expression (1) and under the assumption (2′), i. e., the condition that the curve defined by the ideal (z, x) is ν-fold for the surface F, we shall consider the order of consecutive appearance of the three cases 1, 2 and 3. We have just shown in the proof of the first part of the Theorem that, independently of the different cases, the curve of the surface F_1 given by the ideal (z_1, x_1) is ν-fold. If case 3 occurs infinitely many times our assertion is proved since the curve given by the ideal (z_1, x_1) is in that case the transform of the ν-fold curve given by the ideal (z, x). Now, case 2 cannot take place infinitely many times *consecutively* since by (3. 2) the exponents of the successive parameters x_i decrease strictly. It is also clear from (1. 2) that case 2 or case 3 are the only ones which can occur immediately after case 2. Therefore it remains only to show that case 1 cannot appear infinitely many times consecutively. This is seen as follows: if case 1 occurs i times consecutively (see (4. 1)) the values of the parameters are related in the following form; $0 < v(x_i) = v(x)$ and $0 < v(y_i) = v(y) - iv(x)$. Since v is of rank one $v(x)$ and $v(y)$ may be supposed to be real numbers. Hence $i < v(y)/v(x)$. This completes the proof of the Theorem.

4. It is well known by construction of examples that there exist infinite sequences of infinitely near points on an algebraic surface all of the same multipilicty which are not trivial sequences; see [4] or [6], p. 15. Recently Derwidué[5] has made a new attempt at a purely geometric proof of the theorem of reduction of singularities of an algebraic surface by using Cremona transformations of the ambient 3-space. The proof makes use of a statement concerning the behaviour of the base points of the polar curves and it is asserted that from that statement the following result attributed to B. Levi follows as a simple corollary: an infinite sequence of infinitely near singular points on an algebraic surface is necessarily a trivial sequence. As we have just shown above this assertion is neither true nor does it represent B. Levi's theorem. The procedure follows the classical approach of eliminating the isolated singularities and further reducing the multiple curves but fails to show the true fact that the alternate process of: a) reduction of singular curves, b) elimination of the isolated singularities introduced by a), a_1) reduction of the singular curves introduced by b), etc. is necessarily finite.

HARVARD UNIVERSITY.

REFERENCES.

[1] M. Noether, "Ueber die singulären Werthsysteme einer algebraischen Function und die singulären Punkte einer algebraischen Curve," *Mathematische Annalen*, vol. 9 (1876), pp. 166-182.

[2] C. Segre, "Sulla scomposizione dei punti singolari delle superficie algebriche," *Annali di Matematica*, serie II, vol. 25 (1897), pp. 1-53.

[3] B. Levi, "Sulla riduzione delle singolarita puntuali delle superficie algebriche dello spazio ordinario per trasformazioni quadratiche," *Annali di Matematica*, serie II, vol. 26 (1897), pp. 219-253.

[4] B. Levi, "Intorno alla composizione dei punti generici delle linee singolari delle superficie algebriche," *Annali di Matematica*, serie III, vol. 2 (1899), pp. 127-138.

[5] Enriques-Chisini, *Lezioni sulla teoria geometrica delle equazioni e delle funzioni algebriche*, vol. 2, Zanichelli, 1915-1924.

[6] O. Zariski, "Algebraic surfaces," *Ergebnisse der Mathematik*, vol. 3 (1935), pp. 1-198.

[7] O. Zariski, "Reduction of the singularities of algebraic three dimensional varieties," *Annals of Mathematics*, vol. 45 (1944), pp. 472-542.

[5] L. Derwidué, "Résolution des singularités d'une surface algébrique au moyen de transformations crémoniennes," *Bull. Soc. Roy. Sci. Liège*, vol. 16 (1947), pp. 275-289.

SOME EXAMPLES IN THE THEORY OF SINGULAR BOUNDARY VALUE PROBLEMS.*

By Philip Hartman.**

1. The theorems. Let $q(t)$, where $0 \leqq t < \infty$, be a real-valued, continuous function and λ be a real parameter. Let $N(T, \lambda)$ denote the number of zeros on $0 < t < T$ of a solution $x = x(t) = x(t, \lambda) \not\equiv 0$ of the differential equation

$$x'' + (q + \lambda)x = 0. \tag{1}$$

Thus, up to an additive correction -1, 0 or $+1$ (depending on T), the number $N(T, \lambda)$ is independent of the particular solution $x = x(t)$ of (1) determining it. For example, if $q(t) \equiv 0$, then, as $T \to \infty$,

$$N(T, \lambda) = O(1) \text{ if } \lambda \leqq 0; \qquad N(T, \lambda) = \lambda^{\frac{1}{2}} T/\pi + O(1) \text{ if } \lambda > 0. \tag{2}$$

The asymptotic behavior, as $T \to \infty$, of $N(T, \lambda)$ can depend in a very complicated manner on λ, even for simple functions $q(t)$. To illustrate this, let $q(t)$ be a periodic function, say $q(t) = \cos t$; the solution of the problem of the asymptotic behavior of $N(T, \lambda)$ furnishes the solution of the problem of the determination of the λ-values for which (1) has periodic or half-periodic solutions. Thus, in general, one cannot expect a solution as simple as (2). Nevertheless, it seems surprising that the situation can be as pathological as indicated by the following theorem:

(*) *Let $\psi = \psi(T)$ be a positive, continuous, non-decreasing function for $0 < T < \infty$. Then there exist real-valued, continuous functions $q = q(t)$ for $0 \leqq t < \infty$ realizing each of the following situations, as $T \to \infty$:*

$$N(T, \lambda = O(1) \text{ if } \lambda < 0; \qquad N(T, \lambda) = \psi(T) + O(1) \text{ if } \lambda > 0, \tag{3}$$

$$N(T, \lambda) = \psi(T) + O(1) \text{ for all } \lambda, \tag{4}$$

$$N(T, \lambda) = \psi(T) + O(1) \text{ if } \lambda < 0; \qquad N(T, \lambda) = 2\psi(T) + O(1) \text{ if } \lambda > 0. \tag{5}$$

It is curious that ψ can tend slowly or rapidly to ∞ with T, while the

* Received January 23, 1951.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

right-hand sides of the asymptotic formulae in (3), (4), (5) are essentially independent of λ. The contrast furnished by (2) is indeed great.

Remark on $\lambda = 0$. In the examples, constructed below for the cases (3) and (5) of (*), the number $\lambda = 0$ satisfies the second relation in (3) and (5), respectively; that is, as $T \to \infty$,

(3 bis) $$N(T, 0) = \psi(T) + O(1),$$

(5 bis) $$N(T, 0) = 2\psi(T) + O(1).$$

But it turns out that any situation consistent with Sturm's comparison theorem can be realized. Thus, if $\phi(T)$ is any positive, continuous, non-decreasing function for $0 < T < \infty$ satisfying $\phi(T_2) - \phi(T_1) \leqq \psi(T_2) - \psi(T_1)$ for $0 < T_1 < T_2 < \infty$, then examples $q(t)$ realizing (3) or (5), respectively, can be chosen so that

(3_0) $$N(T, 0) = \phi(T) + O(T)$$

or

(5_0) $$N(T, 0) = \phi(T) + \psi(T) + O(T).$$

A modification of the proof of (*) yields:

(**) *Let S be a closed set on the λ-axis. There exist real-valued, continuous functions $q = q(t)$ on $0 \leqq t < \infty$ such that, for every pair of numbers λ, μ, the difference $N(T, \mu) - N(T, \lambda)$ is unbounded or bounded, as $T \to \infty$, according as the open interval (λ, μ) contains points of S or the closed interval $[\lambda, \mu]$ does not contain points of S.*

Remark. If S is unbounded from below, then $N(T, \lambda) \to \infty$, as $T \to \infty$, for every λ. If S is bounded from below, then examples $q(t)$, proving (**), can be chosen so that each of the following alternatives is realized: as $T \to \infty$, $N(T, \lambda) \to \infty$ for every λ or $N(T, \lambda) = O(1)$ if λ is less than every number in S.

The proofs of (*) and (**) will make it clear that the "pathology" in the examples is not associated with the local smoothness of $q(t)$. In fact, (*) and (**) remain true if the phrase "continuous functions $q(t)$" is replaced by "functions $q(t)$ of class C^∞."

The theorem (*) furnishes the solution to the problem, suggested to me by Professor Wintner, of characterizing the monotone functions $\psi(T)$ corresponding to which there exist differential equations (1) satisfying, as $T \to \infty$, $N(T, \lambda) = O(1)$ if $\lambda < 0$ and $N(T, \lambda) = o(\psi(T))$ if $\lambda > 0$; cf. (3). The theorems (*) and (**) also make it possible to answer some questions, raised by him, concerning the possible structure of the spectra of differential operators (e. g., the question whether the tacit assumption, made by various writers, that the spectrum in the completely continuous case clusters only at ∞, and not at $-\infty$ as well, is not a mistaken one). The answers to these questions are contained in some of the corollaries of the following section.

2. Consequences for the spectral theory of (1). The differential equation (1), for a fixed λ, is said to be *oscillatory* or *non-oscillatory* according as every solution of (1) does or does not have an infinity of zeros. The differential equation (1) is said to be of *limit-circle* or *limit-point* type according as (1) does or does not have two linearly independent solutions of class $L^2(0,\infty)$; the type of (1) is independent of λ; [9], p. 238. In the limit-point case, (1) and a linear homogeneous boundary condition at $t=0$,

$$x(0)\cos\alpha - x'(0)\sin\alpha = 0, \qquad (0 \leqq \alpha < \pi), \tag{6}$$

determine a self-adjoint boundary value problem in $L^2(0,\infty)$; [9]. Let S_α denote the spectrum of this problem and let S' denote the set of (finite) cluster points of S_α. The set S', "the essential spectrum," is independent of α; [9], p. 251. The characterization [2] of S_α and S' in terms of $N(T,\lambda)$, together with (*), (**), gives some curious consequences for the theory of the spectra of differential operators.

No examples are known for which (1) is of limit-point type and S_α contains a point spectrum which clusters at $-\infty$. However, (*) implies the following:

COROLLARY 1. *There exist real-valued, continuous functions $q=q(t)$ on $0 \leqq t < \infty$ such that* (1) *is of limit-point type and S_α (for every α) is a pure point spectrum clustering at, and only at, both ∞ and $-\infty$.*

The spectrum S_α has no finite cluster if and only if the Green kernel, belonging to S_α and a λ (real or non-real) in the complement of S_α, is completely continuous. The known examples of completely continuous Green kernels are the cases where (1) is non-oscillatory for every λ; correspondingly S_α clusters only at ∞. The Green kernels cannot have the property of complete continuity if (1) is oscillatory for some λ and non-oscillatory for some λ; [3]. But examples, proving Corollary 1, show that certain Green kernels can have this property when (1) is oscillatory for every λ (and S_α clusters at both $-\infty$ and ∞).

COROLLARY 2. *There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type, is oscillatory for every* λ *and possesses a non-trivial* ($\not\equiv 0$) *solution of class* $L^2(0, \infty)$ *for every* λ. *Furthermore, the "Green kernels" are completely continuous.*

It is known that S_α always cluster at ∞; [5], p. 310. It is also known [11] that if S_α contains a (non-vacuous) continuous spectrum, then the latter is unbounded. However, it is not known whether or not the essential spectrum S' is necessarily unbounded when it is non-vacuous. Nor is it known whether or not S' can contain an isolated point. The latter two questions are answered by

COROLLARY 3. *There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type and* S_α (*for every* α) *is a pure point spectrum clustering at, and only at,* 0 *and* ∞ (*or at, and only at,* 0 *and both* $-\infty$ *and* ∞).

If (1) is oscillatory for some λ and non-oscillatory for some other λ, then (1) is of limit-point type ([1]) and the least cluster point of S_α is the greatest lower bound, λ_0, of those λ for which (1) is oscillatory; [3]. The spectrum S_α clusters at $\lambda = \lambda_0$ from the left if and only if (1) is oscillatory for $\lambda = \lambda_0$; [1], p. 698. In the known examples with $-\infty < \lambda_0 < \infty$, the spectrum S_α always clusters at $\lambda = \lambda_0$ from the right. But it can be shown that this need not be the case in general.

COROLLARY 4. *There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type, the essential spectrum* S' *consists of the single point* $\lambda = 0$, *but no interval* $(0, \lambda)$, *where* $\lambda > 0$, *contains an infinite subset of* S_α (*for any fixed* α).

Another consequence of (*) is the negative assertion in

COROLLARY 5. *A necessary, but not sufficient, condition that* (1) *be of limit-circle type is that* (1) *be oscillatory and that*

$$N(T, \lambda_2) - N(T, \lambda_1) = O(1), \text{ as } T \to \infty, \tag{7}$$

for every pair of λ-*values* λ_1, λ_2.

In view of the last corollary and the characterization of essential spectra given in [2], the theorem (**) implies:

COROLLARY 6. *Let* S *be a closed set on the* λ-*axis. There exist real-valued, continuous functions* $q = q(t)$ *on* $0 \leqq t < \infty$ *such that* (1) *is of limit-point type and the essential spectrum* S' *of* (1) *is the given set* S.

For example, if S is a perfect, nowhere dense set on the interval $0 \leqq \lambda \leqq 1$, then there exist differential equations (1) of limit-point type for which the set of cluster points of S_α, for every fixed α, is S and the point ∞ or S and the points ∞ and $-\infty$. (Each of the alternatives can be realized by virtue of the Remark following (**).)

In the case just mentioned, or whenever S is bounded, there cannot exist any continuous spectrum, that is, S_α is a pure point spectrum; [11]. Thus, if S is chosen to be the entire interval $0 \leqq \lambda \leqq 1$, one obtains

COROLLARY 7. *There exist real-valued, continuous functions $q = q(t)$ on $0 \leqq t < \infty$ such that* (1) *is of limit-point type and, for every fixed α, the spectrum S_α is a pure point-spectrum which is dense on $0 \leqq \lambda \leqq 1$ and has ∞ (or ∞ and $-\infty$) as its only cluster points not on the interval $0 \leqq \lambda \leqq 1$.*

It is curious to compare this with the result of [5], p. 660.

3. Proof of Corollaries 1 and 2. Choose $\psi(T) = T$ in (*) and let $q = q(t)$ be such that (4) holds. Then it follows that $N(T, \lambda) \sim T = O(T^2)$, as $T \to \infty$, for any λ. Hence, by [6], (1) is of limit-point type. It also follows from (4) that, if $-\infty < \lambda_1 < \lambda_2 < \infty$, then $N(T, \lambda_2) - N(T, \lambda_1) = O(1)$, as $T \to \infty$. Consequently, the characterization of S_α given in [2] implies that S_α has no finite cluster point, that is, that S' is empty. Hence, S_α is a pure point spectrum. But S_α always cluster at ∞; [5], p. 310. Also, S_α clusters at $-\infty$ whenever (1) is oscillatory for every λ; [5], pp. 313-314. This proves Corollary 1.

Corollary 2 is an immediate consequence of Corollary 1. For if λ is not in S', then (1) possesses a non-trivial ($\not\equiv 0$) solution of class $L^2(0, \infty)$; [4]. The complete continuity of the "Green kernels" follows from the remarks preceding the statement of Corollary 2.

4. Proof of Corollary 3. Choose $\psi(T)$ so that $\psi(\infty) = \infty$, and $q(t)$ so that (3) holds. Then (1) is of limit-point type by virtue of the first part of (3); [1]. That $\lambda = 0$ is a cluster point (and the least cluster point $\geqq -\infty$) of S_α follows from [3]. Also $N(T, \lambda_2) - N(T, \lambda_1)$ is bounded or unbounded, as $T \to \infty$, according as λ_1, λ_2 are or are not of the same sign. Consequently, $\lambda = 0$ is the only finite cluster point of S_α, by [2]. Since S_α always clusters at ∞, the proof of the first part of Corollary 3 is complete.

In order to prove the last (parenthetical) part of the Corollary, let $\psi(T) = T$ and let $q(t)$ be such that (5) holds. Then, as in the last section, (1) is of limit-point type and S_α clusters at $\pm\infty$. As in the last paragraph,

$\lambda = 0$ is the only finite cluster point of S_α. This completes the proof of Corollary 3.

5. Proof of Corollary 4. This corollary is a consequence of the proof of the first part of Corollary 3 and the Remark on $\lambda = 0$ in § 1. Let (3 bis) hold, as well (3). Then it follows that $N(T, \lambda) - N(T, 0) = O(1)$, as $T \to \infty$, if $\lambda > 0$. Thus, by [2], the interval $(0, \lambda)$ contains at most a finite number of points of S_α (for any fixed α).

6. Proof of Corollary 5. That a necessary condition that (1) may be of limit-circle type is that (1) be oscillatory has been proved in [1], p. 698. When (1) is of limit-circle type, a boundary condition (3) at $t = 0$ and a similar condition at $t = \infty$ determine a self-adjoint boundary value problem in $L^2(0, \infty)$ with a pure point spectrum, without a finite cluster point; [9]. The characterization in [2] of the spectra S_α for the case that (1) is of limit-point type clearly has an analogue for the spectra associated with the boundary value problems belonging to a differential equation (1) of limit-circle type. Hence, in the latter case, (7) must hold (otherwise the spectrum will have a finite cluster point λ satisfying $\lambda_1 \leqq \lambda \leqq \lambda_2$). Thus the positive part of Corollary 5 follows. The negative part follows from the proof of Corollary 1.

7. Proof of the case (3) of (*). It can be supposed that $\psi(\infty) = \infty$, otherwise (*) is trivial. For if $\psi(\infty) < \infty$, it is sufficient to choose any $q(t)$ satisfying $q(t) \to -\infty$, as $t \to \infty$. Then $q(t) + \lambda$ is negative for large t, and so a solution $x = x(t) \not\equiv 0$ of (1) has no zeros for sufficiently large t. Thus, $N(T, \lambda) = O(1)$, as $T \to \infty$, for every λ. But this is equivalent to (3), (4) and/or (5) when $\psi(\infty) < \infty$.

In order to prove the case (3) of (*), it will first be shown that there exists on $0 \leqq t < \infty$ a step-function $q^*(t)$ such that if $N(T, \lambda)$ refers to

$$y'' + (q^* + \lambda) y = 0, \tag{8}$$

rather than to (1), then (3) holds. This method for obtaining counter-examples in the theory of the differential equation (1) was introduced in [8]. By a solution of (8) is meant a function $y = y(t)$ which, on $0 \leqq t < \infty$, possesses a continuous first derivative and which, on any open interval where q^* is continuous, posssesses a continuous second derivative satisfying (8).

There is no loss of generality in supposing $\psi(0) = 0$ (for otherwise $\psi(T)$ can be suitably altered in a vicinity of $T = 0$, but the behavior of

$\psi(T)$ near $T=0$ does not enter the hypothesis or assertion of (*)). Let $0=a_0<a_1<\cdots$ denote an (unbounded) sequence of T-values satisfying

(9) $$\psi(a_n)=n \text{ for } n=0,1,\cdots.$$

Define a sequence of numbers $(1<)\ \nu_0<\nu_1<\cdots$ in such a way that $\nu_n\to\infty$, as $n\to\infty$, and

(10) $$a_n<b_n<a_{n+1}, \text{ where } b_n=a_n+\pi\nu_n^{-\frac{1}{2}} \text{ and } n=0,1,\cdots.$$

Introduce the abbreviation

(11) $$\mu_n=\nu_{n+1}^5 \text{ for } n=0,1,\cdots.$$

It will be supposed that $\nu_0,\nu_1,\cdots$ increases so rapidly that

(12) $$\nu_{n+1}=o(\mu_n^{\frac{1}{2}}(a_{n+1}-b_n)), \text{ as } n\to\infty.$$

Define $q^*=q^*(t)$ for $0\leqq t<\infty$ as follows:

$$q^*(t)=\nu_n \text{ on } a_n\leqq t<b_n; \qquad q^*(t)=-\mu_n \text{ on } b_n\leqq t<a_{n+1}$$

for $n=0,1,\cdots$. Thus a solution of (8) has the form

(13) $$y(t)=\alpha_n\cos[(\nu_n+\lambda)^{\frac{1}{2}}(t-a_n)]+\beta_n\sin[(\nu_n+\lambda)^{\frac{1}{2}}(t-a_n)],$$

(14) $$y(t)=A_n\exp[(\mu_n-\lambda)^{\frac{1}{2}}(t-b_n)]+B_n\exp[-(\mu_n-\lambda)^{\frac{1}{2}}(t-b_n)],$$

on the respective intervals $a_n<t<b_n$, $b_n<t<a_{n+1}$, for sufficiently large n. Such a solution is determined for large t-values by fixing, say, α_K, β_K for a sufficiently large K and determining the α_n,β_n and the A_n,B_n, for $n\geqq K$, by the conditions that

(15) $$y(t-0)=y(t+0) \text{ and } y'(t-0)=y'(t+0)$$

hold at the points $t=b_K, a_{K+1}, b_{K+1}, a_{K+2},\cdots$.

8. In the sequel, the following abbreviation will be used: If $c_1,c_2,\cdots$ and $d_1,d_2,\cdots$ are two sequences, the symbol $c_n=O(d_n)$, as $n\to\infty$, will signify, as usual, that there exist a constant M such that $|c_n|\leqq M|d_n|$ for all sufficiently large n. But the symbol $c_n=O_M(d_n)$ will signify $|c_n|\leqq M|d_n|$ for the specified constant M and for the specified index n (rather than for some constant M and all sufficiently large n).

Let $\lambda\neq 0$ be fixed. It will be shown that there exists a positive integer $K=K(\lambda)$, so large that $\nu_K+\lambda>0$ and that, for all $n\geqq K$,

(16) $$\sin(\nu_n+\lambda)^{\frac{1}{2}}(b_n-a_n)=-\pi\lambda/2\nu_n+O_1(1/\nu_n^2),$$

(17) $(\nu_n+\lambda)^{\frac{1}{2}}(\mu_n-\lambda)^{-\frac{1}{2}}=(\nu_n/\mu_n)^{\frac{1}{2}}(1+O_{|\lambda|}(1/\nu_n))=O_2(1/\nu_{n+1}{}^2)$,

(18) $(\mu_n-\lambda)^{\frac{1}{2}}(\nu_{n+1}+\lambda)^{-\frac{1}{2}}=\nu_{n+1}{}^2(1+O_{|\lambda|}(1/\nu_{n+1}))$,

and that if $0\leqq h,\ j\leqq 2+|\lambda|$, then for all $n\geqq K$

(19) $(1+O_h(1/\nu_n))(1+O_j(1/\nu_n))^{\pm 1}=1+O_{h+j+1}(1/\nu_n)=O_2(1)$,

and that if C is a fixed constant, say

(20) $$C=18/\pi\,|\lambda|,$$

then for all $n\geqq K$

(21) $$(1+O_C(1/\nu_n))(1+O_C(1/\nu_n))^{-1}\exp[-2(\mu_n-\lambda)^{\frac{1}{2}}(a_{n+1}-b_n)]$$
$$=O_1(1/\nu_{n+1}).$$

The definition of b_n in (10) shows that

$$(\nu_n+\lambda)^{\frac{1}{2}}(b_n-a_n)=(\nu_n+\lambda)^{\frac{1}{2}}\pi/\nu_n^{\frac{1}{2}}=\pi(1+\lambda/\nu_n)^{\frac{1}{2}}=\pi(1+\lambda/2\nu_n+\cdots);$$

so that the left-hand side of (16) is $-\pi\lambda/2\nu_n+(\pi\lambda/2\nu_n)^3/3!+\cdots$, as $n\rightarrow\infty$. Clearly, K can be chosen so large that (16) holds for all $n\geqq K$.

The left-hand side of (17) is

$$(\nu_n/\mu_n)^{\frac{1}{2}}(1+\lambda/\nu_n)^{\frac{1}{2}}(1-\lambda/\mu_n)^{-\frac{1}{2}}$$
$$=(\nu_n/\mu_n)^{\frac{1}{2}}(1+\lambda/2\nu_n+\cdots)(1+\lambda/2\mu_n+\cdots).$$

In view of the definition (11) of μ_n, the last expression is

$$(\nu_n/\mu_n)^{\frac{1}{2}}(1+O_{|\lambda|}(1/\nu_n))$$

if n is sufficiently large. Also, $\nu_n/\mu_n=\nu_n/\nu_{n+1}{}^5\leqq 1/\nu_{n+1}{}^4$. Hence (17) holds for all $n\geqq K$ if K is sufficiently large.

Similarly, it is seen that (18) holds for all $n\geqq K$ if K is sufficiently large. It is also clear that if K is sufficiently large, then (19) is valid for all $n\geqq K$.

As to (21), the exponential factor, for large n, does not exceed exp $[-\mu_n^{\frac{1}{2}}(a_{n+1}-b_n)]$ which, according to (12), is $o(\exp[-\nu_{n+1}])$, as $n\rightarrow\infty$. This makes obvious the existence of a K for which (21) holds when $n\geqq K$.

9. It will be shown that if β_K is chosen to be 1, and α_K is chosen to be $O_2(\beta_K/K^2)$, for example, $\alpha_K=0$, then, in the corresponding solution (13), (14) of (8), for $n\geqq K$,

(22_n) $A_n=-\beta_n(\pi\lambda/4\nu_n)(1+O_C(1/\nu_n))$,

(23_n) $B_n=-\beta_n(\pi\lambda/4\nu_n)(1+O_C(1/\nu_n))$

and

$$(24_n) \qquad \alpha_{n+1} = A_n(1 + O_1(1/\nu_{n+1}))\exp[(\mu_{n+1} - \lambda)^{\frac{1}{2}}(a_{n+1} - b_n)],$$

$$(25_n) \qquad \beta_{n+1} = \nu_{n+1}A_n(1 + O_{|\lambda|+2}(1/\nu_{n+1}))\exp[(\mu_{n+1} - \lambda)^{\frac{1}{2}}(a_{n+1} - b_n)].$$

It can be remarked that if (24_n), (25_n) hold, then by (19),

$$(26_n) \qquad \alpha_{n+1} = O_2(\beta_{n+1}/\nu_{n+1}^2).$$

The proof of (22_n)-(25_n) for $n \geqq K$ will be by mathematical induction on n. Since the passage from n to $n+1$ will depend on (15) and (26_n), it is clear from the choice of α_K, β_K that a direct verification of (22_K)-(25_K) need not be made.

Let $n(\geqq K)$ be fixed and suppose that (22_n)-(25_n) hold; in particular, that (26_n) holds. The relations (15), where $t = b_{n+1}$, give, by virtue of (13), (14),

$$A_{n+1} + B_{n+1} = \alpha_{n+1}\cos + \beta_{n+1}\sin,$$

$$A_{n+1} - B_{n+1} = (\nu_{n+1} + \lambda)^{\frac{1}{2}}(\mu_{n+1} - \lambda)^{-\frac{1}{2}}\{-\alpha_{n+1}\sin + \beta_{n+1}\cos\},$$

where the argument of sin, cos is $(\nu_{n+1} + \lambda)^{\frac{1}{2}}(b_{n+1} - a_{n+1})$. By (16), (17) and (26_n), these equations take the form

$$A_{n+1} + B_{n+1} = O_2(\beta_{n+1}/\nu_{n+1}^2) + \beta_{n+1}(-\pi\lambda/2\nu_{n+1} + O_1(1/\nu_{n+1}^2)),$$

$$A_{n+1} - B_{n+1} = O_2(1/\nu_{n+2}^2)\{O_2(\beta_{n+1}/\nu_{n+1}^2) + O_1(\beta_{n+1})\}.$$

Since $1/\nu_{n+2}^2 \leqq 1/\nu_{n+1}^2 < 1$, it follows that

$$A_{n+1} = -\beta_{n+1}(\pi\lambda/4\nu_{n+1}) + \tfrac{1}{2}O_{2+1+4+2}(\beta_{n+1}/\nu_{n+1}^2).$$

Hence, (22_{n+1}) follows from the definition (20) of C. If the last two simultaneous equations are solved for B_{n+1}, it is seen that (23_{n+1}) holds.

The relations (15), where $t = a_{n+2}$, give

$$\alpha_{n+2} = A_{n+1}\exp[\cdot\ \cdot\ \cdot] + B_{n+1}\exp(-[\cdot\ \cdot\ \cdot]),$$

$$\beta_{n+2} = (\mu_{n+1} - \lambda)^{\frac{1}{2}}(\nu_{n+2} + \lambda)^{-\frac{1}{2}}\{A_{n+1}\exp[\cdot\ \cdot\ \cdot] - B_{n+1}\exp(-[\cdot\ \cdot\ \cdot])\},$$

where

$$(27) \qquad [\cdot\ \cdot\ \cdot] = (\mu_{n+1} - \lambda)^{\frac{1}{2}}(a_{n+2} - b_{n+1}).$$

In view of (18), these equations can be written as

$$(28) \qquad \alpha_{n+2} = A_{n+1}\{1 + (B_{n+1}/A_{n+1})\exp(-2[\cdot\ \cdot\ \cdot])\}\exp[\cdot\ \cdot\ \cdot],$$

(29) $$\beta_{n+2} = \nu_{n+2}^2(1 + O_{|\lambda|}(1/\nu_{n+2}))A_{n+1} \times \{1 - (B_{n+1}/A_{n+1})\exp(-2[\cdot\cdot\cdot])\}\exp[\cdot\cdot\cdot].$$

Since (22_{n+1}), (23_{n+1}) have just been verified, it follows that

$$B_{n+1}/A_{n+1} = (1 + O_C(1/\nu_{n+1}))(1 + O_C(1/\nu_{n+1}))^{-1}.$$

Thus (24_{n+1}) is a consequence of (28) and (21), while (25_{n+1}) follows from (29), (21), and (18). This completes the induction and so (22_n)-(25_n) hold for all $n \geqq K$.

10. Without affecting the preceding considerations, it can be supposed that K has been chosen so large that

(30) $$1 + O_h(1/\nu_n) > \tfrac{1}{2} > 0 \text{ for } n \geqq K \text{ if } h = 1,\ C \text{ or } |\lambda| + 2.$$

It follows from (22_n), (23_n) that

(31) $$A_n \sim B_n, \text{ as } n \to \infty,$$

and that A_n, B_n have the same sign at each $n \geqq K$. Also, (22_{n+1}) and (25_n) show that

(32) $$A_{n+1} = -(\pi\lambda/4\nu_{n+1})A_n(1 + O_C(1/\nu_{n+1}))(1 + O_{|\lambda|+2}(1/\nu_{n+1}))\exp[\cdot\cdot\cdot],$$

where $[\cdot\cdot\cdot]$ is given by (27).

Since A_n and B_n, at each $n \geqq K$, have the same sign, it follows from (14) that $y(t)$ has no zero on $b_n \leqq t \leqq a_{n+1}$. If $\lambda > 0$, then A_n and A_{n+1} are of opposite signs, by (30), (32). It follows, therefore, that $y(t)$ has an odd number of zeros on $a_{n+1} < t < b_{n+1}$. But (13) and (10) show that $y(t)$ has on this last interval at most $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}^{\frac{1}{2}}$ zeros. Since $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}^{\frac{1}{2}} \to 2$, as $n \to \infty$, it follows that $y(t)$ has, for large n, exactly one zero on $a_{n+1} < t < b_{n+1}$. Thus the second part of (3) is a consequence of the definition, (9), of a_n.

If $\lambda < 0$, then A_n and A_{n+1} are of the same sign, and so $y(t)$ has an even number of zeros on $a_{n+1} < t < b_{n+1}$. But, on this interval, $y(t)$ has at most $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}^{\frac{1}{2}}$ zeros. Since $1 + (\nu_{n+1} + \lambda)^{\frac{1}{2}}/\nu_{n+1}^{\frac{1}{2}} < 2$ when $\lambda < 0$, it follows that $y(t)$ has no zero on $a_{n+1} < t < b_{n+1}$. Thus the first part of (3) is proved.

11. This proves the case (3) of (*) if $q(t)$ is allowed to be a step-function, $q^*(t)$. It will be shown that $q^*(t)$ can be modified so as to become a continuous function $q(t)$, for which (3) remains valid.

Since every solution $y = y(t)$ of (8) has an absolutely continuous first derivative, one has $d(y^2 + y'^2) = 2y'(y + y'')\,dt = -2yy'(q^* + \lambda - 1)\,dt$. Thus $|\,d(y^2 + y'^2)\,| \leqq (y^2 + y'^2)(|\,q^*\,| + |\,\lambda\,| + 1)\,dt$; and so $y^2 + y'^2 = O(\exp \int_0^t (|\,q^*(s)\,| + |\,\lambda\,| + 1)\,ds)$, as $t \to \infty$. In particular,

$$(33) \qquad |\,y(t)\,| = O(\Omega(t)), \text{ as } t \to \infty,$$

where $\Omega(t) = \exp \frac{1}{2} \int_0^t (|\,q^*(s)\,| + s)\,ds$. The point in the inequality (33) is that $y(t)$, being a solution of (8), depends on λ but the majorant $\Omega(t)$ is independent of λ.

Starting with $q^*(t)$ construct a continuous function $q(t)$ on $0 \leqq t < \infty$ by letting $q(t) = q^*(t)$ except on small intervals, say $(b_0, b_0 + \epsilon_0)$, $(a_1 - \epsilon_1, a_1)$, $(b_1, b_1 + \epsilon_1)$, $(a_2 - \epsilon_2, a_2), \cdots$, where $q(t)$ is linear. Let

$$(34) \qquad \epsilon(t) = q(t) - q^*(t).$$

It is clear that if the numbers $\epsilon_0, \epsilon_1, \cdots$ tend to 0 sufficiently rapidly, then

$$(35) \qquad \int^{\infty} |\,\epsilon(t)\,|\,\Omega^2(t)\,dt < \infty.$$

Let $y = y_1(t) = y_1(t, \lambda)$ and $y = y_2(t) = y_2(t, \lambda)$ be a pair of solutions of (8) satisfying the Wronskian condition $y'_1 y_2 - y_1 y'_2 \equiv 1$. If (1) is written as $x'' + (q^* + \lambda)x = (q^* - q)x$, it is seen that every solution $x = x(t) = x(t, \lambda)$ of (1), and its derivative, can be written in the form

$$(36) \qquad x = p_2 y_1 - p_1 y_2 \text{ and } x' = p_2 y'_1 - p_1 y'_2$$

where, for $j = 1, 2$,

$$p_j = p_j(t, \lambda) = \gamma_j + \int_0^t (q^*(s) - q(s))\,y_j(s)\,x(s)\,ds,$$

and γ_1. γ_2 are constants. The results of [10], pp. 261-268, show that (33) and (35) imply that $c_j = \lim p_j(t, \lambda)$, as $t \to \infty$, exist for $j = 1, 2$; furthermore, $c_1 = c_2 = 0$ only if $x(t) \equiv 0$. (In [10], q and q^* are assumed to be continuous, but it is clear that the discontinuities of q^* do not affect the arguments there.)

Let $x(t) = x(t, \lambda) \not\equiv 0$; so that at least one the numbers c_1, c_2 is not zero. For the sake of concreteness, let $c_1 \neq 0$. Then the Wronskian condition on y_1, y_2 and (36) imply $xy'_1 - x'y_1 = -p_1 \to -c_1$, as $t \to \infty$. Consequently, for large t-values, x and y_1 do not vanish simultaneously and $(\arctan y_1/x)'$ does not change sign; and so, the zeros of y_1 and x separate each other; cf. also [7]. Thus (3) holds for (1), since it holds for (8). This completes the proof of the case (3) of (*).

12. Proof of the case (4) of (*). This part of (*) is proved in a manner similar to that of case (3). The main differences will be indicated. In (10), let the numbers b_n be defined by

$$b_n = a_n + 3\pi/2\nu_n^{\frac{1}{2}} \qquad \text{for } n = 0, 1, \cdots. \tag{37}$$

Then (16) can be replaced by

$$\sin(\nu_n + \lambda)^{\frac{1}{2}}(b_n - a_n) = -1 + O_M(1/\nu_n^2), \tag{38}$$

where, say, $M = M(\lambda) = 5\lambda^2(>(3\pi\lambda/4)^2/2!)$. By a suitable choice of the integer K and the constant D (depending on λ), it can be shown that the corresponding equation (8) has a solution (13), (14), where, for $n \geqq K$,

$$A_n = -(\beta_n/2)(1 + O_D(1/\nu_n)), \quad B_n = -(\beta_n/2)(1 + O_D(1/\nu_n)),$$

while (24_n), (25_n) hold.

In this case, A_n and A_{n+1} are of opposite sign for all λ. Thus, for the differential equation (8), the relation (4) can be verified as in § 10. The passage from (8) to (1) is the same as in § 11.

13. Proof of the case (5) of (*). This can be proved by another choice of b_n, namely,

$$b_n = a_n + 2\pi/\nu_n^{\frac{1}{2}} \qquad \text{for } n = 0, 1, \cdots. \tag{39}$$

Then (16) becomes

$$\sin(\nu_n + \lambda)^{\frac{1}{2}}(b_n - a_n) = \pi\lambda/\nu_n + O_1(1/\nu_n^2). \tag{40}$$

By the procedure of § 9, it can be shown that the corresponding equation (8) has a solution (13), (14), where, for $n \geqq K$,

$$A_n = \beta_n(\pi\lambda/2\nu_n)(1 + O_C(1/\nu_n)), \quad B_n = \beta_n(\pi\lambda/2\nu_n)(1 + O_C(1/\nu_n)),$$

while (24_n), (25_n) hold. Clearly, the proof can be finished by the procedures of §§ 10-11.

This completes the proof of (*).

14. Proof of (3 bis) and (5 bis). Return to the proof of the case (3) in § 9. It will be shown that if $\lambda = 0$ and K is a suitably chosen integer, then (8) has a solution (13), (14) where, for $n \geqq K$,

$$(41_n) \qquad \begin{aligned} A_n &= -\tfrac{1}{2}\beta_n(\nu_n^{-2} + \nu_{n+1}^{-2})(1 + O_8(1/\nu_n)), \\ B_n &= -\tfrac{1}{2}\beta_n(\nu_n^{-2} - \nu_{n+1}^{-2})(1 + O_8(1/\nu_n)), \end{aligned}$$

and

$$(42_n) \qquad \begin{aligned} \alpha_{n+1} &= A_n(1 + O_1(1/\nu_{n+1}))\exp[\mu_n^{\frac{1}{2}}(a_{n+1} - b_n)], \\ \beta_{n+1} &= \nu_{n+1}^2 A_n(1 + O_1(1/\nu_{n+1}))\exp[\mu_n^{\frac{1}{2}}(a_{n+1} - b_n)]. \end{aligned}$$

The proof of these formulae, by induction, will be similar to the proof of (22_n)-(25_n). The relations (42_n) imply, if n is sufficiently large,

$$(43_n) \qquad \alpha_{n+1} = (\beta_{n+1}/\nu_{n+1}^2)(1 + O_4(1/\nu_{n+1})).$$

Let K be a positive integer. to be fixed below. Let α_K, β_K be chosen so that (43_{K-1}) holds.

Assume (41_n) and (42_n). The equations for A_{n+1}, B_{n+1}, in the formula lines following (26) become for $\lambda = 0$

$$A_{n+1} + B_{n+1} = -\alpha_{n+1}, \qquad A_{n+1} - B_{n+1} = -\nu_{n+2}^{-2}\beta_{n+1}.$$

Thus, by (43_n),

$$A_{n+1} = -\tfrac{1}{2}\beta_{n+1}(\nu_{n+2}^{-2} + \nu_{n+2}^{-2}) + O_2(\beta_{n+1}/\nu_{n+1}^3),$$

$$B_{n+1} = -\tfrac{1}{2}\beta_{n+1}(\nu_{n+2}^{-2} - \nu_{n+2}^{-2}) + O_2(\beta_{n+1}/\nu_{n+1}^3),$$

and so (41_{n+1}) holds.

The equations for α_{n+2}, β_{n+2} become identical with those preceding (27), where $\lambda = 0$. These can be written in the form

$$\alpha_{n+2} = A_{n+1}\{1 + (B_{n+1}/A_{n+1})\exp(-2[\cdot\ \cdot\ \cdot])\}\exp[\cdot\ \cdot\ \cdot],$$

$$\beta_{n+2} = \nu_{n+2}^2 A_{n+1}\{1 - (B_{n+1}/A_{n+1})\exp(-2[\cdot\ \cdot\ \cdot])\}\exp[\cdot\ \cdot\ \cdot],$$

where $[\cdot\ \cdot\ \cdot]$ is given by (27) when $\lambda = 0$. By (41_{n+1}), which has just been verified, $B_{n+1}/A_{n+1} = O_1(1)(1 + O_8(1/\nu_{n+1}))(1 + O_8(1/\nu_{n+1}))^{-1}$. Clearly, if K is fixed sufficientiy large, then (42_{n+1}) holds.

The proof can now be completed as in § 10. The case $\lambda = 0$ of (5) is similar.

15. Proof of the Remark on (3_0) and (5_0). Let a_n, b_n, ν_n, μ_n have the same significance as in (9), (10), (11), (12). Define $\delta_0, \delta_1, \cdots$ by placing $\delta_n = \pm 1/\nu_n^{\frac{1}{2}}$, where the $\pm$ is a function of n to be given below. Let $q^*(t)$ be defined, for $0 \leqq t < \infty$, as follows:

$$q^*(t) = \nu_n + \delta_n \text{ on } a_n \leqq t < b_n; \qquad q^*(t) = -\mu_n + \delta_n \text{ on } b_n \leqq t < a_{n+1}$$

for $n = 0, 1, \cdots$. Let (13'), (14') denote the equation which result if λ is replaced by $\lambda + \delta_n$ in (13), (14). Then the solutions of (8), for large t, are given by (13'), (14'). It is clear from §§ 8-9 that if $\lambda \neq 0$, then (8) has a solution (13'), (14') where, for large n, the coefficients satisfy the equations which result if $\lambda + \delta_n$, $\lambda + \delta_{n+1}$ are written in place of λ in (22_n)-(23_n), (24_n)-(25_n), respectively. Thus, by the arguments of § 10, the asymptotic formulae (3) hold if $N(T, \lambda)$ refers to (8).

It remains to consider (3_0). It will be shown that (8), where $\lambda = 0$, has a solution (13'). (14') with coefficients satisfying

$$(44_n) \qquad \begin{aligned} A_n &= -\beta_n(\pi\delta_n/4\nu_n)(1 + O_M(1/\nu_n^{\frac{1}{2}})), \\ B_n &= -\beta_n(\pi\delta_n/4\nu_n)(1 + O_M(1/\nu_n^{\frac{1}{2}})) \end{aligned}$$

and

$$(45_n) \qquad \begin{aligned} \alpha_{n+1} &= A_n(1 + O_3(1/\nu_{n+1}))\exp[(\mu_{n+1} - \delta_{n+1})^{\frac{1}{2}}(a_{n+1} - b_n)], \\ \beta_{n+1} &= \nu_{n+1}{}^2 A_n(1 + O_3(1/\nu_{n+1}))\exp[(\mu_{n+1} - \delta_{n+1})^{\frac{1}{2}}(a_{n+1} - b_n)], \end{aligned}$$

for large n, where $M = 18/\pi$.

The equations (45_n) imply (26_n), if n is sufficiently large. Let K be a positive integer, to be fixed below. Let α_K, β_K be chosen so that (26_{K-1}) holds and assume that, for a fixed $n(\geqq K)$, the relations (44_n), (45_n) hold.

The equations for A_{n+1}, B_{n+1} become those in the formula lines following (26_n), where λ must be replaced by δ_{n+1}. Since (16) and (17) hold for large n, with $\lambda = \delta_n$, the equations A_{n+1}, B_{n+1} become identical with those in the second pair of formula lines following (26_n), with the modification that λ is δ_{n+1}. Hence, the definition of δ_n shows that (44_{n+1}) holds (if K is sufficiently large).

Since (18) can be improved, for $\lambda = \delta_n$ and large n, to

$$(\mu_n - \delta_n)^{\frac{1}{2}}(\nu_{n+1} + \delta_{n+1})^{-\frac{1}{2}} = \nu_{n+1}{}^2(1 + O_1(1/\nu_{n+1})),$$

the deduction of (24_n), (25_n) above shows that (45_{n+1}) is a consequence of (44_{n+1}) if n is sufficiently large.

Let K be so large that (8), where $\lambda = 0$, has a solution $y = y(t)$ for which (44_n), (45_n) hold for $n \geqq K$. The arguments of § 10 show that, for large n, $y(t)$ has one or no zero for $a_n \leqq t < a_{n+1}$ according as the $+$ or $-$ sign holds in the definition of δ_n. It is clear that, for a given $\phi(T)$, the $\pm$ signs (as function of n) can be chosen so that (3_0) holds if $N(T, 0)$ refers to (8). Since the passage from (8) to (1) is the same as in § 11, the case (3_0) is proved.

Clearly, the case (5_0) is proved similarly with the choice $b_n = a_n + 2\pi/\nu_n^{\frac{1}{2}}$.

16. Proof of (**). It can be supposed that the given set S is not empty. Otherwise it is sufficient to choose $q(t)$ so that $q(t) \rightarrow -\infty$, as $t \rightarrow \infty$. Then $N(T, \mu) - N(T, \lambda) = O(1)$, as $T \rightarrow \infty$, for $-\infty < \lambda < \mu < \infty$.

Clearly, § 11 implies that it is sufficient to prove the existence of a function $q^*(t)$, $0 \leqq t < \infty$, which is a step-function with discontinuities clustering only at $t = \infty$ and for which the assertion (**) holds if (1) is replaced by (8).

Let $\tau^1, \tau^2, \cdots$ be a sequence of points of the λ-set S (allowing repetitions) with the property that every point of S is either a point or a cluster point of the sequence. Let $\tau_1, \tau_2, \cdots$ denote the sequence $\tau^1; \tau^1, \tau^1, \tau^2, \tau^2;$ $\tau^1, \tau^1, \tau^1, \tau^2, \tau^2, \tau^2, \tau^3, \tau^3, \tau^3; \tau^1, \cdots$; so that, for a given k, there exist arbitrarily large N and M such that $\tau_n = \tau^k$ for $n = N, N+1, \cdots, M$.

Let $\psi(t)$ be a continuous, monotone function satisfying $\psi(0) = 0$ and $\psi(\infty) = \infty$, and let (9), (10), (11) and (12) hold. In addition, it can be supposed that $\nu_0, \nu_1, \cdots$ increases so rapidly that, as $n \rightarrow \infty$,

$$\sigma_n{}^3 = o(\nu_n), \text{ where } \sigma_n = n + \max(|\tau_1|, \cdots, |\tau_n|).$$

Define $q^* = q^*(t)$, for $0 \leqq t < \infty$, as follows:

$$q^*(t) = \nu_n - \tau_n \text{ for } a_n \leqq t < b_n \text{ and } q^*(t) = -\mu_n - \tau_n \text{ for } b_n \leqq t < a_{n+1},$$

where $n = 0, 1, \cdots$. Corresponding to (13), (14), a solution of (8) has, for large n, the form

$$(46) \quad y(t) = \alpha_n \cos[(\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(t - a_n)] + \beta_n \sin[(\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(t - a_n)]$$

(47) $$y(t) = A_n \exp[(\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(t - b_n)] + B_n \exp[-(\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(t - b_n)]$$

for $a_n \leqq t < b_n$, $b_n \leqq t < a_{n+1}$, respectively.

Let $\epsilon > 0$ and λ be fixed. A slight modification of the procedures in § 8 shows that there exists a positive integer $K = K(\lambda, \epsilon)$ so large that for $n \geqq K$ one has $\nu_n + \lambda - \tau_n > 0$ and

(48) $$\sin(\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(b_n - a_n) = -\pi(\lambda - \tau_n)/2\nu_n + O_1(1/\nu_n^2),$$

(49) $$(\nu_n + \lambda - \tau_n)^{\frac{1}{2}}(\mu_n - \lambda + \tau_n)^{-\frac{1}{2}} = (\nu_n/\mu_n)^{\frac{1}{2}}(1 + O_1(\sigma_n/\nu_n)) = O_2(1/\nu_{n+1}^2),$$

(50) $$(\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(\nu_{n+1} + \lambda - \tau_n)^{-\frac{1}{2}} = \nu_{n+1}^2(1 + O_1(\sigma_{n+1}/\nu_{n+1})),$$

(51) $$(1 + O_1(1/\nu_n))(1 + O_1(\sigma_n/\nu_n)) = 1 + O_1((2 + \sigma_n)/\nu_n) = O_2(1),$$

(52) $$(1 + O_h(1/\nu_n)(1 + O_j(1/\nu_n)) = (1 + O_{h+j+1}(1/\nu_n)), 0 \leqq h, j \leqq 9,$$

and, in addition, for

(53) $$C = 18/\pi\epsilon,$$

one has

(54) $$(1 + O_C(1/\nu_n))(1 + O_C(1/\nu_n))^{-1} \exp(-2\{\cdot\ \cdot\ \cdot\}) = O_1(1/\nu_{n+1}),$$

where

(55) $$\{\cdot\ \cdot\ \cdot\} = (\mu_n - \lambda + \tau_n)^{\frac{1}{2}}(a_{n+1} - b_n).$$

These relations and the arguments of § 9 show that if $N \geqq K(\lambda, \epsilon)$ and

(56) $$|\lambda - \tau_n| \geqq \epsilon > 0 \text{ for } n = N, N + 1, \cdots, M,$$

then (8) has a solution $y = y(t)$ for $a_N \leqq t \leqq b_{M+1}$, given by (46), (47), where $\alpha_N = O_2(\beta_N/\nu_N^2)$ and, for $n = N, N + 1, \cdots, M$,

(57_n) $$A_n = -(\beta_n \pi(\lambda - \tau_n)/4\nu_n)(1 + O_C(1/\nu_n)),$$

(58_n) $$B_n = -(\beta_n \pi(\lambda - \tau_n)/4\nu_n)(1 + O_C(1/\nu_n))$$

and

(59_n) $$\alpha_{n+1} = A_n(1 + O_1(1/\nu_{n+1})) \exp\{\cdot\ \cdot\ \cdot\},$$

(60_n) $$\beta_{n+1} = \nu_{n+1}^2 A_n(1 + O_1((2 + \sigma_{n+1})/\nu_{n+1})) \exp\{\cdot\ \cdot\ \cdot\}.$$

If K is so large that $|O_1((2 + \sigma_n)/\nu_n)| < \frac{1}{2}$, $|O_C(1/\nu_n)| < \frac{1}{2}$ and $1 + (\nu_{n+1} + \lambda - \tau_{n+1})^{\frac{1}{2}}/\nu_{n+1}^{\frac{1}{2}} \leqq 5/2 < 3$ whenever $n \geqq K$, then the methods of

§ 10 show that (57_n)-(60_n) imply that $y = y(t)$ has exactly one or no zero on the intervals $b_n \leqq t \leqq b_{n+1}$ according as $\lambda > \tau_n$ or $\lambda < \tau_n$, where $n = N, N+1, \cdots, M$.

Let $\lambda < \mu$ and suppose that there exists an $\epsilon > 0$ such that, for all sufficiently large n,

$$(61) \qquad | \lambda - \tau_n | \geqq \epsilon > 0 \text{ and } | \mu - \tau_n | \geqq \epsilon > 0.$$

Then $N(T, \mu) - N(T, \lambda)$ is unbounded or bounded according as there is or is not at least one value of n (hence infinitely many) for which $\lambda < \tau_n < \mu$. This proves the assertion (**), for (8), in the case in which S has no point on the closed interval $[\lambda, \mu]$.

If the open interval (λ, μ) contains points of S, the structure of the sequence $\tau_1, \tau_2, \cdots$ shows that there exist an $\epsilon > 0$ and a sequence of increasing integers $N_1 < M_1 < N_2 < M_2 < \cdots$ such that $M_n - N_n \to \infty$, as $n \to \infty$, and $\lambda + \epsilon < \tau_n < \mu - \epsilon$ if $n = N_k, N_k + 1, \cdots, M_k$, where $k = 1, 2, \cdots$. Thus, (56) holds for $N = N_k$ and $M = M_k$. Hence (8) has on the interval $b_N \leqq t \leqq b_M$, where $N = N_k$ and $M = M_k$, a solution with no zeros; while if λ is replaced by μ, the resulting differential equation (8) has a solution with exactly $M - N$ zeros. Thus the contribution of $b_N \leqq t \leqq b_M$ to $N(T, \mu) - N(T, \lambda)$, for $T \geqq b_M$, is at least $(M - N) - 2$. Consequently, $M_n - N_n \to \infty$, as $n \to \infty$, implies that $N(T, \mu) - N(T, \lambda) \to \infty$, as $T \to \infty$. This completes the proof of (**).

17. Proof of the Remark following (**). If S is bounded from below, the above proof of (**) gives an example in which (1) is non-oscillatory for λ less than every number in S. In order to obtain an example in which (1) is oscillatory for every λ, it is sufficient to repeat the above construction, with b_n in (10) replaced by $b_n = a_n + 2\pi / \nu_n^{\frac{1}{2}}$.

APPENDIX.*

In the paper "Oscillatory and non-oscillatory differential equations" by Wintner and the author, this Journal, vol. 71 (1949), p. 646, the question is raised as to whether or not the condition

$$(62) \qquad q(t) \text{ tends monotonously to } \infty, \text{ as } t \to \infty,$$

implies that the essential spectrum of (1) is the entire λ-axis, when (2) is of limit-point type. It was proved *loc. cit.* that the answer is in the affirmative

* Recieved May 5, 1951.

when the growth of q is sufficiently smooth. The object of this Appendix is to adapt the procedures employed above to prove that, in general, the answer is in the negative.

There exist on $0 \leqq t < \infty$ *continuous functions* $q(t)$ *having the properties that* (62) *holds, that* (1) *is of limit-point type, and that the spectrum* S_α *is a pure point spectrum clustering at, and only at,* ∞ *and* $-\infty$.

This assertion can be considered as a refinement of Corollaries 1 and 2 in § 2. The considerations of § 11 show that it is sufficient to construct a step-function $q^*(t)$ having all of the desired properties, except that of continuity. It turns out that the example in the Appendix of [6], pp. 211-212, is of this type. In this particular example, the function q (or q^*) satisfies

$$\int^{\infty} dt/q^{\frac{1}{2}} < \infty.$$

It will remain an open question whether or not this holds for every example.

The function $q^*(t)$ in the Appendix of [6] is defined on $0 \leqq t < \infty$ by

$$q^*(t) = \nu_n \text{ if } a_n \leqq t < a_{n+1} \tag{63}$$

and

$$a_1 = 0 \text{ and } a_{n+1} - a_n = 2\pi/\nu_n^{\frac{1}{2}} \text{ for } n = 1, 2, \cdots, \tag{64}$$

finally

$$\nu_n = n^2; \tag{65}$$

so that $a_n \to \infty$, while $q^*(t)$ satisfies (62). This function makes (8) of limit-point type, by [6], pp. 211-212. It will be shown that, for every λ, (8) has a solution which vanishes exactly twice on $a_n \leqq t < a_{n+1}$ for large n; in particular, $N(T, \lambda) - N(T, \mu) = O(1)$, as $T \to \infty$, for $-\infty < \lambda < \mu < \infty$. This implies, by [2], the statement concerning S_α.

Every solution of (8) is for large t of the form

$$y(t) = \alpha_n \cos (\nu_n + \lambda)^{\frac{1}{2}}(t - a_n) + \beta_n \sin (\nu_n + \lambda)^{\frac{1}{2}}(t - a_n), \tag{66}$$

where some pair of constants α_K, β_K can be chosen arbitrarily and $\alpha_{K+1}, \beta_{K+1}, \alpha_{K+2}, \cdots$ are chosen so as to satisfy (15). The relations between α_n, β_n and $\alpha_{n+1}, \beta_{n+1}$ are given by

$$\begin{aligned} \alpha_{n+1} &= \alpha_n \cos[\cdots] + \beta_n \sin[\cdots], \\ \beta_{n+1} &= (\nu_n + \lambda)^{\frac{1}{2}}(\nu_{n+1} + \lambda)^{-\frac{1}{2}}\{-\alpha_n \sin[\cdots] + \beta_n \cos[\cdots]\}, \end{aligned} \tag{67}$$

where the argument $[\cdot\ \cdot\ \cdot]$ of sin and cos is, by (64), $2\pi(\nu_n+\lambda)^{\frac{1}{2}}\nu_n^{-\frac{1}{2}}$, which is of the form $2\pi(1+\lambda/2\nu_n+\cdot\ \cdot\ \cdot)$, if λ fixed and $n\rightarrow\infty$. Consequently, if n is sufficiently large, $\cos[\cdot\ \cdot\ \cdot]=1+O_M(\lambda^2/\nu_n^2)$ and $\sin[\cdot\ \cdot\ \cdot]=O_M(\lambda/\nu_n)$, where $M=\pi^2$, and $(\nu_n+\lambda)^{\frac{1}{2}}(\nu_{n+1}+\lambda)^{-\frac{1}{2}}\leqq 1-1/2\nu_n^{\frac{1}{2}}$, by (65). Hence, (67) becomes

$$\begin{aligned} &\alpha_{n+1}=\alpha_n(1+O_M(\lambda^2/\nu_n^2))+O_M(\lambda\beta_n/\nu_n), \\ (68)\qquad & \\ &|\,\beta_{n+1}\,|\leqq(1-1/2\nu_n^{\frac{1}{2}})(O_M(\lambda\alpha_n/\nu_n)+|\,\beta_n\,|). \end{aligned}$$

If $\alpha_n\neq 0$, put $c_n=\beta_n/\alpha_n$. It will be shown, by induction, that if the integer $K=K(\lambda)$ is sufficiently large, $\alpha_K\neq 0$ and $c_K=O_M(\lambda/\nu_{K-1})$, then $\alpha_{n+1}\neq 0$, and the relation

$$(69_n)\qquad c_{n+1}=O_M(\lambda\sum_{k=K-1}^{n}1/\nu_k)\leqq C,\ \text{where}\ C=C(K,\lambda)$$

is a constant (cf. (65)), holds for $n=K-1,K,\cdot\ \cdot\ \cdot$. Assume $\alpha_n\neq 0$ and that (69_{n-1}) holds, where $n(\geqq K)$ is fixed. Then the first equation of (68) gives

$$(70_n)\qquad \alpha_{n+1}=\alpha_n(1+O_{2MC}(\lambda/\nu_n)),\ \text{where}\ 1+O_{2MC}(\lambda/\nu_n)>0,$$

if K is sufficiently large. In particular, $\alpha_{n+1}\neq 0$. The second equation of (68) shows that

$$|\,c_{n+1}\,|\leqq(1-1/2\nu_n^{\frac{1}{2}})(1+O_{3MC}(\lambda/\nu_n))(M\,|\,\lambda\,|/\nu_n+|\,c_n\,|).$$

Since the product of the first two factors on the right does not exceed 1 if K is sufficiently large, (69_n) follows from (69_{n-1}). This completes the induction.

Consequently, (70_n) holds for $n=K,K+1,\cdot\ \cdot\ \cdot$. Since $y(a_n)=\alpha_n$ by (66), it follows that $y(a_n)$ and $y(a_{n+1})$ are of the same sign, so that $y(t)$ vanishes an even number of times on $a_n<t<a_{n+1}$. That $y(t)$, for large t, vanishes exactly twice on $a_n\leqq t<a_{n+1}$, follows by the arguments of § 10. This completes the proof of the italicized assertion.

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "Differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[2] ———, "A characterization of the spectra of one-dimensional wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 915-920.

[3] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[4] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[5] ——— and A. Wintner, "On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

[6] ——— and A. Wintner, "A criterion for the non-degeneracy of the wave equation," *ibid.*, vol. 71 (1949), pp. 206-213.

[7] W. Leighton, "On self-adjoint differential equations of second order," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 656-657.

[8] H. Milloux, "Sur l'équation differentielle $x'' + A(t)x = 0$," *Prace Matematyczno-Fizyczne*, vol. 41 (1934), pp. 39-54.

[9] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[10] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

[11] ———, "On Dirac's theory of continuous spectra," *Physical Review*, vol. 73 (1948), pp. 781-785.

BOOLEAN ALGEBRAS WITH OPERATORS.*

By Bjarni Jónsson and Alfred Tarski.

PART II **

Section 4.

Representation Theorems for Relation Algebras.

Relation algebras are abstract algebraic systems characterized by means of a number of simple postulates which prove to be satisfied if we take the elements of the algebra to be binary relations, and the fundamental operations of this algebra to be set-theoretical operations of addition and multiplication together with relative multiplication and conversion. The relation algebra actually formed by binary relations and the operations just mentioned will be referred to as proper relation algebras. The natural representation problem for relation algebras is the problem whether every relation algebra is isomorphic to a proper relation algebra. It has recently been shown that in general the solution of this problem is negative.[14] On the other hand, we shall see in Theorem 4. 22 that every relation algebra has at least a "weak" natural representation in which all the operations except the Boolean multiplication have their natural meaning. In some further theorems, e. g., 4. 29 and 4. 32, we shall obtain a positive solution of the natural representation problem for certain classes of relation algebras which, however, are of a rather special nature.

When studying abstract relation alegbras it is useful to bear in mind that various notions of the general theory of these algebras take on a familiar meaning and various results can easily be anticipated when applied to proper relation algebras.

Definition 4. 1. *An algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

* Received March 13, 1950.

** Part I appeared in vol. 73 (1951), pp. 891-939, of this Journal.

[14] Cf. Lyndon [1]. The numbers in brackets refer to the bibliography at the beginning of Part I of this paper.

(where $+$, $\cdot$, and ; are operations on A^2 to A, $^\smile$ is an operation on A to A, and 0, 1, *and* 1' *are elements of A) is called a relation algebra if the following conditions are satisfied:*

(i) $\langle A, +, 0, \cdot, 1\rangle$ *is a Boolean algebra.*

(ii) $(x\,;y);z = x\,;(y\,;z)$ *for any* $x, y, z \in A$.

(iii) $1'\,;x = x = x\,;1'$ *for every* $x \in A$.

(iv) *The formulas* $(x\,;y)\cdot z = 0$, $(x^\smile;z)\cdot y = 0$, *and* $(z;y^\smile)\cdot x = 0$ *are equivalent for any* $x, y, z \in A$.

The operation ; is referred to as relative multiplication, the operation $^\smile$ as conversion, and the element 1' *as the identity element.*[15]

In view of condition (ii) of this definition we shall in general, when speaking of relation algebras, omit parentheses in expressions like

$$(x;y);z \text{ and } x;(y;z).$$

Condition 4.1(iv) plays a fundamental role in the theory of relation algebras. It is useful to notice in this connection the following

THEOREM 4.2. *In Definition* 4.1 *condition* (iv) *can be replaced by the following one:*

(iv′) *Given any element $a \in A$, the functions f and g defined by the formulas*

$$f(x) = a;x \text{ and } g(x) = a^\smile;x \text{ for every } x \in A$$

are conjugate, and so are the functions f' and g' defined by the formulas

$$f'(x) = x;a \text{ and } g'(x) = x;a^\smile \text{ for every } x \in A.$$

Proof. By 1.11 and 4.1.

We shall not develop here either the arithmetic of relation algebras or the proper algebraic theory of these algebras (the study of isomorphisms, homomorphisms, subalgebras, etc.)—except insofar as it is relevant for the main purposes of our discussion. Some arithmetical consequences of 4.1 are stated in the next theorem.

THEOREM 4.3. *For any relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^\smile\rangle$$

---

[15] The axiom system 4.1(i)-(iv) is equivalent to the one given in Jónsson-Tarski [2]. For the proof of the equivalence of the two systems see Chin-Tarski [1], Theorem 2.2; for the relation of these systems to the axiom system in Tarski [2] see Chin-Tarski [1], footnote 10.

we have:

(i) *If $x \varepsilon A$, I is an arbitrary set, and the elements $y_i \varepsilon A$ with $i \varepsilon I$ are such that $\sum_{i \varepsilon I} y_i \varepsilon A$, then*

$$\sum_{i \varepsilon I} (x ; y_i) \varepsilon A, \quad \sum_{i \varepsilon I} (y_i ; x) \varepsilon A,$$
$$x ; (\sum_{i \varepsilon I} y_i) = \sum_{i \varepsilon I} (x ; y_i), \text{ and } (\sum_{i \varepsilon I} y_i) ; x = \sum_{i \varepsilon I} (y_i ; x).$$

(ii) *If $x, x', y, y' \varepsilon A$, $x \leq x'$, and $y \leq y'$, then $x ; y \leq x' ; y'$.*

(iii) *$x ; 0 = 0 = 0 ; x$ for every $x \varepsilon A$.*

(iv) *$x \leq x ; 1$ and $x \leq 1 ; x$ for every $x \varepsilon A$.*

(v) *$(x + y)^{\cup} = x^{\cup} + y^{\cup}$, $(x \cdot y)^{\cup} = x^{\cup} \cdot y^{\cup}$, and $(x ; y)^{\cup} = y^{\cup} ; x^{\cup}$ for any $x, y \varepsilon A$.*

(vi) *If $x, y \varepsilon A$ and $x \leq y$, then $x^{\cup} \leq y^{\cup}$.*

(vii) *$x^{\cup\cup} = x$ and $x^{-\cup} = x^{\cup -}$ for every $x \varepsilon A$.*

(viii) *$(x ; y) \cdot z \leq x ; x^{\cup} ; z$ for any $x, y, z \varepsilon A$.*

(ix) *$x \leq x ; x^{\cup} ; x$ for every $x \varepsilon A$.*

(x) *$x^{\cup} ; (x ; y)^{-} \leq y^{-}$ for any $x, y \varepsilon A$.*

(xi) *$0^{\cup} = 0$, $1^{\cup} = 1$, and $(1')^{\cup} = 1'$.*

(xii) *If $x \varepsilon A$ is an atom, then $x^{\cup}$ is an atom.*

Proof. The proof of parts (i)-(xi) of this theorem can be found elsewhere [16]; (xii) obviously follows from (vi), (vii), and (xi).

THEOREM 4. 4. *Every relation algebra is a normal Boolean algebra with operators.*

Proof. By 2. 13, 4. 1, and 4. 3(i) (iii) (v) (xi).

Some further arithmetical notions applying to relation algebras are introduced in the following.

DEFINITION 4. 5. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\cup} \rangle$$

be a relation algebra.

(i) *The element $(1')^{-}$ is called the diversity element and is denoted by $0'$. An element $x \varepsilon A$ is referred to as*

(ii) *an equivalence element if $x ; x \leq x$ and $x^{\cup} \leq x$,*

[16] See Chin-Tarski [1], §§ 1 and 2.

(iii) *a functional element if* $x^{\smile} ; x \leq 1'$,

(iv) *an ideal element if* $x = 1 ; x ; 1$.

The reasons why we have chosen the terms introduced in 4. 5(ii)(iii)(iv) will become clear in our further discussion (see Theorem 4. 24 and remarks preceding Definition 4. 8).

THEOREM 4. 6. *Given a relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

we have

(i) $(0')^{\smile} = 0'$.

(ii) $0' ; 0' ; 0' ; 0' = 0' ; 0'$.

(iii) *For every* $x \in A$ *the following three conditions are equivalent:* x *is an equivalence element,* $x ; x = x = x^{\smile}$, *and* $x^{\smile} ; x = x$.

(iv) *If* $x, y \in A$, x *is a functional element, and* $y \leq x$, *then* y *is a functional element.*

(v) *If* $x, y, z \in A$, *and* x *is a functional element, then*

$$x ; (y \cdot z) = (x ; y) \cdot (x ; z) \text{ and } (y \cdot z) ; x^{\smile} = (y ; x^{\smile}) \cdot (z ; x^{\smile}).$$

(vi) *If* $x \in A$ *and* $x \leq 1'$, *then* x *is both an equivalence element and a functional element.*

(vii) *If* $x, y \in A$, x *is an atom, and* y *is a functional element, then* $x ; y = 0$ *or* $x ; y$ *is an atom.*

(viii) *If* $x, y \in A$ *are ideal elements, then* $x + y$, $x \cdot y$, *and* x^{-} *are ideal elements.*

(ix) *If* $x, y \in A$ *are ideal elements, then* $x ; y = x \cdot y$ *and* $x^{\smile} = x$.

(x) *If* $x, y, z \in A$ *and* x *is an ideal element, then*

$$x \cdot (y ; z) = (x \cdot y) ; (x \cdot z).$$

(xi) 0 *and* 1 *are ideal elements.*

(xii) $1 ; x ; 1$ *is an ideal element for every* $x \in A$.

(xiii) $x \in A$ *is an ideal element if, and only if,* x *and* x^{-} *are equivalence elements.*

Proof. The proof of all parts of this theorem, except (vii), can be found elsewhere.[17] To prove (vii), suppose x is an atom and y is a func-

[17] See Chin-Tarski [1], § 3.

tional element. Consider any element $z \neq 0$ such that $z \leq x ; y$. Then $(x ; y) \cdot z \neq 0$ and hence, by 4.1(iv), $(z ; y^{\smile}) \cdot x \neq 0$. Since x is an atom, this implies that $x \leq z ; y^{\smile}$. Hence, by 4.1(ii)(iii), 4.3(ii), and 4.5(iii),

$$x ; y \leq z ; y^{\smile} ; y \leq z ; 1' = z,$$

and therefore $z = x ; y$. Since this is true for every $z \neq 0$ with $z \leq x ; y$, we obtain the conclusion of (vii) at once.

We shall now establish some facts belonging to the proper algebraic theory of relation algebras.

THEOREM 4.7. (i) *A homomorphic image of a relation algebra is again a relation algebra.*

(ii) *A cardinal product of relation algebras is again a relation algebra.*

Proof. The corresponding theorem for Boolean algebras is well known. Hence and from the form of conditions (ii), (iii), and (iv) of 4.1 we see directly that 4.7(ii) holds; to obtain 4.7(i) we notice in addition that, by 1.15(i)(iii) and 4.2, condition 4.1(iv) can be equivalently replaced by a system of equations.

By an *ideal* in the relation algebra

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, {}^{\smile} \rangle$$

we understand any non-empty set $J \subseteq A$ satisfying the conditions:

(i) if $x, y \in J$, then $x + y \in J$;

(ii) if $x \in J$ and $y \in A$, then $x \cdot y \in J$, $x ; y \in J$, and $y ; x \in J$.[18]

The connections between homomorphisms and ideals in relation algebras prove to be entirely analogous to those in Boolean algebras or arbitrary rings. Moreover, it turns out that the discussion of ideals in relation algebras reduces entirely to the discussion of ideals in Boolean algebras. For a one-to-one correspondence which preserves the inclusion relation can be established between ideals in the relation algebra $\mathfrak{A}$ and those in the Boolean algebra

$$\mathfrak{J} = \langle I, +, 0, \cdot, 1 \rangle$$

where I is the set of all ideal elements $a \in A$. (The system $\mathfrak{J}$ is clearly a Boolean algebra by 4.6(viii)(xi).) In particular, given any $a \in I$, the set of all elements $x \in A$ with $x \leq a$ is a principal ideal in $\mathfrak{A}$ which corresponds

[18] See Chin-Tarski [1], § 3 (remarks following Definition 3.23).

to the principal ideal in $\mathfrak{J}$ consisting of all elements $x \varepsilon I$ with $x \leq a$; if a is not in I, the set of all $x \varepsilon A$ with $x \leq a$ is not an ideal in $\mathfrak{A}$. Without using the notion of an ideal explicitly, we give a few theorems, 4.9-4.14, which are suggested by the above remarks and by the knowledge of analogous results applying to Boolean algebras.

DEFINITION 4.8. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

be a relation algebra. Let a be any element in A, and B be the set of all elements $x \varepsilon A$ such that $x \leq a$. Then the system

$$\langle B, +, 0, \cdot, a, ;, a \cdot 1', {}^{\smile} \rangle$$

will be denoted by $\mathfrak{A}(a)$.

THEOREM 4.9. *If*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

is a relation algebra and $a \varepsilon A$ is an ideal element, then $\mathfrak{A}(a)$ is a relation algebra and the function ϕ on A defined by the formula

$$\phi(x) = a \cdot x \text{ for every } x \varepsilon A$$

maps $\mathfrak{A}$ homomorphically onto $\mathfrak{A}(a)$.

Proof. It is well known that ϕ maps homomorphically the Boolean algebra $\langle A, +, 0, \cdot, 1 \rangle$ onto $\langle B, +, 0, \cdot, a \rangle$. Moreover, by 4.6(x) we have

$$\phi(x ; y) = \phi(x) ; \phi(y) \text{ for any } x, y \varepsilon A,$$

while, by 4.3(v) and 4.6(ix),

$$\phi(x^{\smile}) = [\phi(x)]^{\smile} \text{ for every } x \varepsilon A.$$

Since, in addition, $\phi(1') = a \cdot 1'$, we conclude by 4.8 that ϕ maps $\mathfrak{A}$ homomorphically onto $\mathfrak{A}(a)$. Hence, by 4.7(i), $\mathfrak{A}(a)$ is a relation algebra.

Notice that a necessary and sufficient condition for $A(a)$ to be a relation algebra is that a be an equivalence element (and not necessarily an ideal element).

THEOREM 4.10. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

be a relation algebra in which $0 \neq 1$. *Then the following conditions are equivalent:*

(i) $\mathfrak{A}$ *is simple.*

(ii) $\mathfrak{A}$ *has no ideal elements different from* 0 *and* 1.

(iii) *For every* $x \varepsilon A$, $x \neq 0$ *implies* $1 ; x ; 1 = 1$.

(iv) *For any* $x, y \varepsilon A$, $x ; 1 ; y = 0$ *implies that* $x = 0$ *or* $y = 0$.[19]

Proof. By 4.9, (i) implies (ii). By 4.3(iv) and 4.6(xii), (ii) implies (iii). Now assume that (iii) holds. Let ϕ be a homomorphism mapping $\mathfrak{A}$ onto another algebra

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, ;, 1', ^\smile \rangle$$

and suppose B contains at least two different elements. By 4.7(i), $\mathfrak{B}$ is also a relation algebra. If $x, y \varepsilon A$ and $\phi(x) = \phi(y)$, and if we put

$$z = x \cdot y^- + y \cdot x^-,$$

then $\phi(z) = 0$, and hence, by 4.3(iii), $\phi(1 ; z ; 1) = 0$. Hence $\phi(1 ; z ; 1) \neq 1 = \phi(1)$, and therefore $1 ; z ; 1 \neq 1$. It follows by (iii) that $z = 0$ whence $x = y$. Thus, ϕ is an isomorphism. Consequently, (iii) implies (i), and conditions (i)-(iii) are equivalent.

Suppose now again that (iii) holds, and consider any elements $x, y \varepsilon A$ such that $x ; 1 ; y = 0$. By 4.3(iii) we obtain $1 ; x ; 1 ; y = 0$. Hence, by (iii), if $x \neq 0$, then $1 ; y = 0$ and therefore, by 4.3(iv), $y = 0$. Thus (iii) implies (iv). Finally, assume (iv) to hold. From 4.3(x), with x and y replaced by $1 ; x$ and 1, respectively we obtain $(1 ; x)^\smile ; (1 ; x ; 1)^- \leq 0$, and therefore, by 4.3(v)(xi), $x^\smile ; 1 ; (1 ; x ; 1)^- = 0$. Hence, by applying (iv) and with the help of 4.3(vii)(xi), we conclude that, if $x \neq 0$, then $(1 ; x ; 1)^- = 0$ and $1 ; x ; 1 = 1$. Thus, (iv) implies (iii), and the proof is complete.

THEOREM 4.11. *Every relation algebra which is a subalgebra of a simple relation algebra is itself simple.*

Proof. By 4.10(i)(iii).

THEOREM 4.12. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', ^\smile \rangle$$

be a relation algebra. Let I be an arbitrary set, and, for every i ε I, let

$$\mathfrak{B}_i = \langle B_i, +_i, 0_i, \cdot_i, 1_i, ;_i, 1'_i, ^{\smile_i} \rangle$$

[19] This is a joint result of J. C. C. McKinsey and A. Tarski; see Jónsson-Tarski [2].

be a relation algebra. In order that $\mathfrak{A}$ *be isomorphic to the cardinal product of the algebras* $\mathfrak{B}_i$ *it is necessary and sufficient that there exist elements* $a_i \varepsilon A$ *satisfying the following conditions:*

(i) $$\sum_{i \varepsilon I} a_i = 1.$$

(ii) *If* $i, j \varepsilon I$ *and* $i \neq j$, *then* $a_i \cdot a_j = 0$.

(iii) $\sum_{i \varepsilon I} (x_i \cdot a_i) \varepsilon A$ *for any elements* $x_i \varepsilon A$ *correlated with* $i \varepsilon I$.

(iv) a_i *is an ideal element for every* $i \varepsilon I$.

(v) $\mathfrak{A}(a_i) \cong \mathfrak{B}_i$ *for every* $i \varepsilon I$.

(*Condition* (iii) *is automatically satisfied in case the set* I *is finite.*)

Proof. Let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

be the cardinal product of all the algebras $\mathfrak{B}_i$ with $i \varepsilon I$. B is then the set of all functions ϕ on I such that $\phi(i) \varepsilon B_i$ for every $i \varepsilon I$. Given $i \varepsilon I$, let b_i be the unique function $\phi \varepsilon B$ such that

$$\phi(i) = 1_i, \text{ and } \phi(j) = 0_j \textit{ for every } j \varepsilon I,\ j \neq i.$$

As is easily seen, using 4. 3(iii)(iv) and 4. 5(iv), the elements $b_i \varepsilon B$ thus defined satisfy conditions (i)-(v) of our theorem (with 'A' and 'a_i' changed to 'B' and 'b_i'). Hence, if $\mathfrak{A}$ is isomorphic to $\mathfrak{B}$, there are also elements $a_i \varepsilon A$ satisfying the same conditions.

Assume now, conversely, that there are elements $a_i \varepsilon A$ satisfying conditions (i)-(v). For any given $x \varepsilon A$ let $\psi(x)$ be the only function ϕ on I such that

$$\phi(i) = a_i \cdot x \text{ for every } i \varepsilon I.$$

Then, by (i)-(iii), ψ maps the set A in one-to-one way onto the set of all functions ϕ on I such that $\phi(i) \leq a_i$ for every $i \varepsilon I$. Hence, with the help of (iv), 4. 8, and 4. 9, we conclude that ψ maps $\mathfrak{A}$ isomorphically onto the direct product of the algebras $\mathfrak{A}(a_i)$ with $i \varepsilon I$. Consequently, by (v), $\mathfrak{A}$ is isomorphic to the direct product $\mathfrak{B}$ of the algebras $\mathfrak{B}_i$ with $i \varepsilon I$. Thus, our theorem holds in both directions.

THEOREM 4. 13. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

as well as $\mathfrak{B}$ *and* $\mathfrak{C}$ *be relation algebras. In order that* $\mathfrak{A} \cong \mathfrak{B} \times \mathfrak{C}$ *it is necessary and sufficient that there be an ideal element* $a \varepsilon A$ *such that*

$$\mathfrak{A}(a) \cong \mathfrak{B} \text{ and } \mathfrak{A}(a^-) \cong \mathfrak{C}.$$

Proof. By 4.6(viii) and 4.12.

THEOREM 4.14. *A relation algebra is simple if, and only if, it is indecomposable.*

Proof. By 4.8, 4.10, and 4.13 a relation algebra which is not simple cannot be indecomposable. The converse is known to hold for every algebra.

THEOREM 4.15. *For every relation algebra $\mathfrak{A}$ there exist simple relation algebras $\mathfrak{B}_i$ correlated with elements i of a set I such that $\mathfrak{A}$ is isomorphic to a subalgebra of the cardinal product of the algebras $\mathfrak{B}_i$ and that, for every $i \in I$, $\mathfrak{B}_i$ is a homomorphic image of $\mathfrak{A}$.*

Proof. If we replace in 4.15 "relation algebras" by "algebras" and "simple" by "indecomposable," we obtain a statement which holds for arbitrary algebras and is a direct consequence of a result known from the literature.[20] Hence, by restricting ourselves to relation algebras and applying 4.7(i) and 4.14, we obtain 4.15 at once.

It may be mentioned that relation algebras and specifically simple relation algebras are closely related to cylindric algebras discussed in the preceding section. In fact, if

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

is a relation algebra, and if we put

$$C_0(x) = 1 ; x \text{ and } C_1(x) = x ; 1 \text{ for every } x \in A,$$

then

$$\mathfrak{A}' = \langle A, +, 0, \cdot, 1, C_0, C_1 \rangle$$

proves to be a generalized cylindric algebra. Moreover, if $\mathfrak{A}$ is simple, then $\mathfrak{A}'$ is also simple, and hence $\mathfrak{A}'$ is a cylindric algebra in the sense of 3.15 (compare remarks following 3.18).

Besides simple algebras, a more special class of relation algebras—in fact, that of integral algebras defined below in 4.16—will be involved in a part of our further discussion (see Theorems 5.10-5.12 in the following section). An even more important role will be played by still another class of relation algebras—in fact, by algebras in which every atom is a functional element; we do not introduce any special term to denote such algebras. Some general properties of these two classes of relation algebras will be established in the following theorems.

[20] See Birkhoff [2].

DEFINITION 4.16. *A relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

is said to be an integral relation algebra (or a relation algebra without zero divisors) if $0 \neq 1$ *and if, for any* $x, y \,\varepsilon\, A$, *the formula* $x ; y = 0$ *implies that* $x = 0$ *or* $y = 0$.

THEOREM 4.17. *For every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

in which $0 \neq 1$ *the following conditions are equivalent:*

(i) $\mathfrak{A}$ *is an integral relation algebra.*

(ii) *For every* $x \,\varepsilon\, A$, $x \neq 0$ *implies* $x ; 1 = 1$.

(iii) *Every functional element* $x \,\varepsilon\, A$ *such that* $x \neq 0$ *is an atom.*

(iv) $1'$ *is an atom.*

Proof. Assume (i) to hold. By 4.3(x) with $y = 1$ we have $x^{\smile} ; (x ; 1)^{-} = 0$ for every $x \,\varepsilon\, A$. Hence, by (i) and 4.16, $x^{\smile} = 0$ or $x ; 1 = 1$. If $x \neq 0$, then, by 4.3(vii)(xi), $x^{\smile} \neq 0$ and therefore $x ; 1 = 1$. Thus, (i) implies (ii).

Suppose that (ii) holds. By 4.3(viii), with $y = 1$ and $z = 1'$, and 4.1(iii), we obtain for every $y \,\varepsilon\, A$:

(1) If $y \neq 0$, then $1' = (y ; 1) \cdot 1' \leq y ; y^{\smile} ; 1' = y ; y^{\smile}$.

Let now $x \neq 0$ be a functional element and let y be any element such that $y \leqq x$ and $y \neq 0$. By 4.1(iii). 4.3(ii)(vi), 4.5(iv), and (1) we have

$$x = 1' ; x \leq y ; y^{\smile} ; x \leq y ; x^{\smile} ; x \leq y ; 1' = y.$$

Hence $y = x$. Thus, x is an atom, and (ii) implies (iii).

By 4.6(vi), $1'$ is a functional element. Also, $1' \neq 0$ since otherwise we should have, by 4.1(iii) and 4.3(iii).

$$1 = 1 ; 1' = 1 ; 0 = 0,$$

contrary to the hypothesis of the theorem. Hence, (iii) implies (iv).

Now assume that (iv) holds, and consider any elements $x, y \,\varepsilon\, A$ for which $x ; y = 0$. By (iv), either $(y ; 1) \cdot 1' = 1'$ or $(y ; 1) \cdot 1' = 0$. In the first case we have $1' \leqq y ; 1$, and hence, by 4.1(ii)(iii) and 4.3(ii)(iii),

$$x = x ; 1' \leq x ; y ; 1 = 0 ; 1 = 0.$$

In the second case we obtain, by 4.1(iv), $(y^{\smile} ; 1') \cdot 1 = 0$, and hence, by

4. 1 (iii) and 4. 3 (vii) (xi), $y = 0$. Consequently, by 4. 16, (iv) implies (i). Thus, conditions (i)-(iv) are equivalent.

From 4. 3 (v) (vii) (xi) it is easily seen that formula $x \mathbin{;} 1 = 1$ in 4. 17 (ii) can be replaced by $1 \mathbin{;} x = 1$.

THEOREM 4. 18. (i) *Every relation algebra which is a subalgebra of an integral relation algebra is itself integral.*

(ii) *Every integral relation algebra is simple.*

Proof. (i) immediately follows from 4. 16. To obtain (ii) we notice that, by 4. 3 (ii) (iv) and 4. 16, condition 4. 17 (i) directly implies 4. 10 (iv). (Similarly, 4. 17 (ii) directly implies 4. 10 (iii).)

THEOREM 4. 19. *For every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

the following conditions are equivalent:

(i) *Every atom $x \in A$ is a functional element.*

(ii) *If $x, y \in A$ are atoms, then $x \mathbin{;} y = 0$ or $x \mathbin{;} y$ is an atom.*

(iii) *If $x \in A$ is an atom, then $x^{\smile} \mathbin{;} x$ is an atom.*

Proof. Observe that, for every $x \in A$, $(x^{\smile} \mathbin{;} x) \cdot 1' = 0$ implies $x = 0$ by 4. 1 (iii) (iv). Hence

(1) $$(x^{\smile} \mathbin{;} x) \cdot 1' \neq 0 \text{ whenever } x \neq 0.$$

By 4. 6 (vii), condition (i) implies (ii). Assume now (ii) to hold. If $x \neq 0$ is an atom, then $x^{\smile}$ is an atom by 4. 3 (xii), and therefore $x^{\smile} \mathbin{;} x$ is an atom by (ii) and (1). Thus, (ii) implies (iii). From (1) and 4. 5 (iii) we also see that (iii) implies (i), and the proof is complete.

THEOREM 4. 20. *For every atomistic relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

in which $0 \neq 1$ the following two conditions are equivalent:

(i) *$\mathfrak{A}$ is integral and every atom $x \in A$ is a functional element.*

(ii) *If $x, y \in A$ are atoms, then $x \mathbin{;} y$ is also an atom.*

Proof. By 4. 16 and 4. 19 (i) (ii), condition (i) implies (ii) for every relation algebra $\mathfrak{A}$ (whether atomistic or not). Again by 4. 19 (i) (ii), condition (ii) implies that every atom in $\mathfrak{A}$ is a functional element, and since $\mathfrak{A}$ is atomistic, this condition also implies by 4. 3 (ii) and 4. 16 that $\mathfrak{A}$ is integral. Thus conditions (i) and (ii) are equivalent.

We now turn to the main subject of this section—representation theorems for relation algebras. We begin with the extension theorem:

THEOREM 4.21. *For every relation algebra* $\mathfrak{B}$ *there is a complete and atomistic relation algebra* $\mathfrak{A}$ *which is a perfect extension of* $\mathfrak{B}$. *If* $\mathfrak{B}$ *is simple, or integral, the same applies to* $\mathfrak{A}$.

Proof. By 2.15 and 4.4 there is a complete and atomistic Boolean algebra with operators $\mathfrak{A}$ which is a perfect extension of $\mathfrak{B}$. From 2.11(i), 2.18, and 4.1 it follows that $\mathfrak{A}$ is a relation algebra. The second part of the theorem results from 2.11(ii), 4.10(i) (iii), and 4.17(i) (ii).

THEOREM 4.22. *Every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

is isomorphic to a relation algebra

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \wedge, \odot, V, |, I, {}^{-1} \rangle$$

where A *is a family of binary relations* $R \subseteq V$, V *is an equivalence relation, and* I *is the identity function on the field of* V, *while* $\cup$, $\wedge$, $|$, *and* ${}^{-1}$ *have their usual (set-theoretical) meaning.*

Proof. By 4.21 no loss of generality arises if we restrict ourselves to the case when $\mathfrak{A}$ is atomistic. Let, in this case, U be the set of all atoms of $\mathfrak{A}$. We define a function F on A by means of the formula

(1) $$F(x) = \underset{\langle a,b \rangle}{\mathrm{E}} [a, b \,\varepsilon\, U \text{ and } a \leq x \,;\, b] \text{ for every } x \,\varepsilon\, A.$$

Furthermore, we put

(2) $$\mathsf{A} = F^*(A), \qquad V = F(1), \qquad \text{and } I = F(1').$$

Assume that $F(x) = F(y)$ for any given $x, y \,\varepsilon\, A$. $\mathfrak{A}$ being atomistic, this assumption implies by (1) that

(3) $$x \,;\, b = y \,;\, b \text{ for every } b \,\varepsilon\, U.$$

By 4.1(iii) and 4.3(i), again in view of the atomistic character of $\mathfrak{A}$, we have

(4) $$x = x \,;\, 1' = x \,;\, (\sum_{1' \geq b \,\varepsilon\, U} b) = \sum_{1' \geq b \,\varepsilon\, U} (x \,;\, b),$$

and similarly

(5) $$y = \sum_{1' \geq b \,\varepsilon\, U} (y \,;\, b).$$

Formulas (3)-(5) imply that $x = y$. Thus, by (2),

(6) $$F \text{ maps } A \text{ onto } \mathsf{A} \text{ in one-to-one way.}$$

By (1) and 4.3(i) (iii) we have

(7) $$F(x + y) = F(x) \cup F(y) \quad \text{for any } x, y \in A$$

and

(8) $$F(0) = \Lambda.$$

Consider any $x, y \in A$. If $\langle a, b\rangle \in F(x\,;y)$, then, by (1), $a \leq x\,;(y\,;b)$ and $a, b \in U$. Hence $\mathfrak{A}$ being atomistic, we conclude by 4.3(i) that there is a $c \in U$ such that $a \leq x\,;c$ and $c \leq y\,;b$. Therefore, by (1), $\langle a, c\rangle \in F(x)$, $\langle c, b\rangle \in F(y)$, and consequently $\langle a, b\rangle \in F(x) \mid F(y)$. In a similar way, using (1) and 4.3(ii), we show that the latter formula implies $\langle a, b\rangle \in F(x\,;y)$. Thus,

(9) $$F(x\,;y) = F(x) \mid F(y) \quad \text{for any } x, y \in A.$$

Suppose $x \in A$. If $\langle a, b\rangle \in [F(x)]^{-1}$, i. e., $\langle b, a\rangle \in F(x)$, then by (1), $b \leq x\,;a$ and $a, b \in U$. Therefore $(x\,;a) \cdot b \neq 0$ whence $(x^{\smile};b) \cdot a \neq 0$ by 4.1(iv). Consequently, $a \leq x^{\smile};b$ and $\langle a, b\rangle \in F(x^{\smile})$. Thus

$$[F(x)]^{-1} \subseteq F(x^{\smile}) \quad \text{for every } x \in A.$$

Hence the inclusion in the opposite direction can easily be derived with the help of 4.3(vii), so that finally

(10) $$F(x^{\smile}) = [F(x)]^{-1} \quad \text{for every } x \in A.$$

In view of (6), F has the inverse function F^{-1} which maps A onto A. If we now define the binary operation $\odot$ on elements of A by putting

$$R \odot S = F(F^{-1}(R) \cdot F^{-1}(S)) \quad \text{for } R, S \in \mathsf{A},$$

we clearly have

(11) $$F(x \cdot y) = F(x) \odot F(y) \quad \text{for any } x, y \in A.$$

By (2) and (6)-(11), the functions F maps $\mathfrak{A}$ isomorphically onto

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \odot, V, \mid, I, {}^{-1}\rangle.$$

Hence $\mathfrak{A}'$ is a relation algebra. By 4.3(iv)(xi) we have $V \mid V = V = V^{-1}$; thus, V is an equivalence relation. By (1), (2), and 4.3(iv), the field of V is U. Finally, from (1), (2), and 4.1(iii) we see that I is the set of all couples $\langle a, a\rangle$ with $a \in U$—or, in other words, the identity function on U to U. The proof has thus been completed.

It should be emphasized that, in general, the operation $\odot$ in the algebra $\mathfrak{A}'$ of Theorem 4.22 is not set-theoretical multiplication; i.e., R and S being two relations in A, $R \odot S$ does not necessarily coincide with the intersection $R \cap S$ of R and S. However, the meaning of $\odot$ is unambiguously determined by the fact that, according to 4.1, $\langle \mathsf{A}, \cup, \Lambda, \odot, V \rangle$ is a Boolean algebra; $R \odot S$ is the largest relation in A included in both R and S. As was pointed out at the beginning of this section, the representation theorem 4.22 cannot be improved in the sense that the operation $\odot$ in $\mathfrak{A}'$ cannot be assumed to coincide with set-theoretical multiplication $\cap$. In the remaining part of this section we shall be concerned with those algebras $\mathfrak{A}'$ in which $\odot$ coincides with $\cap$, and we shall prove a few special representation theorems in which such algebras are involved.

Definition 4.23. *A relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

is called a proper relation algebra if A is a family of binary relations, $0 = \Lambda$, and if $R + S = R \cup S$, $R \cdot S = R \cap S$, $R ; S = R \mid S$, and $R^{\smile} = R^{-1}$ for any $R, S \in \mathsf{A}$. $\mathfrak{A}$ is called a proper relation algebra on a set U if, in addition, U is a non-empty set, $1 = U^2$, and $1'$ is the identity function on U.

When referring to proper relation algebras, we shall use the symbol V instead of 1, and the symbol I instead of $1'$.

Theorem 4.24. *Let*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle$$

be a proper relation algebra.

(i) *For a relation $R \in \mathsf{A}$ to be an equivalence element in $\mathfrak{A}$ it is necessary and sufficient that R be an equivalence relation. In particular, V and I are equivalence relations having the same field.*

(ii) *If I is an identity function, then, for a relation $R \in \mathsf{A}$ to be a functional element in $\mathfrak{A}$, it is necessary and sufficient that R be a function (a many-to-one relation).*

Proof. The first part of (i) obviously follows from 4.5(ii) and 4.23. Hence, by 4.6(vi)(xi)(xiii), V and I are equivalence relations. Since, by 4.1(i)(iii), $I \subseteq V \subseteq I \mid V$, we see that V and I have the same field. Finally, (ii) is an obvious consequence of 4.5(iii) and 4.23.

In connection with 4.24 it may be noticed that the notion of an ideal element also assumes a rather simple meaning when applied to proper relation algebras. In fact, $\mathfrak{A}$ being a proper relation algebra with the universal relation V, let U be the field of V. Since, by 4.24(i), V is an equivalence relation, there exists a partition of U under V, i. e., a family K of non-empty mutually exclusive sets X such that

$$U = \bigcup_{X \varepsilon \mathsf{K}} X \text{ and } V = \bigcup_{X \varepsilon \mathsf{K}} X^2.$$

It is now easily seen that a relation $R \varepsilon \mathsf{A}$ is an ideal element in the sense of 4.5(iv) if, and only if, for some family $\mathsf{L} \subseteq \mathsf{K}$,

$$R = \bigcup_{X \varepsilon \mathsf{L}} X^2.$$

THEOREM 4.25. *For*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle$$

to be a proper relation algebra, it is necessary and sufficient that the following conditions be satisfied:

(i) *A is a set-field whose elements are binary relations and V is the universal set of A.*

(ii) *$R \mid S \varepsilon \mathsf{A}$ and $R^{-1} \varepsilon \mathsf{A}$ for any $R, S \varepsilon \mathsf{A}$.*

(iii) *$I \varepsilon \mathsf{A}$, and $R \mid I = R = I \mid R$ for every $R \varepsilon \mathsf{A}$.*

Proof. Obvious, by 4.1 and 4.23.

THEOREM 4.26. *Let A be the family of all binary relations R with $R \subseteq V$, and let*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle.$$

(i) *For $\mathfrak{A}$ to be a proper relation algebra, it is necessary and sufficient that V be an equivalence relation and I be the identity function on the field of V.*

(ii) *For $\mathfrak{A}$ to be a simple proper relation algebra it is necessary and sufficient that, for some non-empty set U, $V = U^2$ and I be the identity function on U (in other words, that $\mathfrak{A}$ be a proper relation algebra on a set U).*

Proof. If $\mathfrak{A}$ is a proper relation algebra, then V is an equivalence relation by 4.24(i). J being the identity function on the field of V, we clearly have $J \mid I = I$ and also, by 4.25(iii), $J \mid I = J$, so that $I = J$. If, conversely, V is an equivalence relation and I is the identity function on the field of V, we easily see (e. g., by 4.25) that $\mathfrak{A}$ is a proper relation algebra.

Assume that $\mathfrak{A}$ is a simple proper relation algebra. By (i), V is an equivalence relation and, U being the field of V, I is the identity function on U. Since $V \neq \Lambda$, the set U is not empty. For any given $a \,\varepsilon\, U$ we denote by $V(a)$ the set of all couples $\langle x, y\rangle \,\varepsilon\, V$ such that $\langle a, x\rangle \,\varepsilon\, V$ and $\langle a, y\rangle \,\varepsilon\, V$. We easily see that

$$V \mid V(a) \mid V = V(a).$$

Since $V(a) \neq \Lambda$, we conclude by 4.10(i) (iii) that $V(a) = V$. Hence, if $x, y \,\varepsilon\, U$, then $\langle x, x\rangle \,\varepsilon\, V(a)$, $\langle y, y\rangle \,\varepsilon\, V(a)$, and therefore $\langle x, y\rangle \,\varepsilon\, V(a) = V$; consequently, $V = U^2$. Thus, the conditions stated in (ii) are necessary for $\mathfrak{A}$ to be simple. If, conversely, $V = U^2$ and I is the identity function on U, then, by (i), $\mathfrak{A}$ is a proper relation algebra. Moreover, we easily check that $V \mid R \mid V = V$ for every $R \subseteq V$ such that $R \neq \Lambda$. Hence, by 4.10(i) (iii), $\mathfrak{A}$ is simple. The proof has thus been completed.

It should be pointed out at this place that, in general the relation I in a proper relation algebra $\mathfrak{A}$ is not always an identity function. Also, even in case $\mathfrak{A}$ is simple, the relation V is not necessarily of the form $V = U^2$, and hence $\mathfrak{A}$ is not necessarily a proper relation algebra on a set U. In this connection the following two theorems deserve attention.

THEOREM 4.27. *Every proper relation algebra*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, \mid, I, ^{-1}\rangle$$

is isomorphic to a proper relation algebra

$$\mathfrak{A}' = \langle \mathsf{A}', \cup, \Lambda, \cap, V', \mid, I', ^{-1}\rangle$$

where I' is the identity function on the field of V'.

Proof. For every x let $x^{\odot} = I^*(\{x\})$ and for every $R \,\varepsilon\, A$ let $F(R)$ be the set of all ordered pairs of the form $\langle x^{\odot}, y^{\odot}\rangle$ with $\langle x, y\rangle \,\varepsilon\, R$. Moreover let

$$\mathsf{A}' = F^*(\mathsf{A}), \ V' = F(V) \text{ and } I' = F(I).$$

It is easy to check that the function F maps $\mathfrak{A}$ isomorphically onto the system

$$\mathfrak{A}' = \langle \mathsf{A}', \cup, \Lambda, \cap, V', \mid, I'\rangle,$$

whence it follows that $\mathfrak{A}'$ is a relation algebra. Finally, I' is clearly an identity function on the field of V'.

THEOREM 4.28. *For a proper relation algebra $\mathfrak{A}$ to be simple it is necessary and sufficient that $\mathfrak{A}$ be isomorphic to a proper relation algebra $\mathfrak{A}'$ on a set U.*

Proof. U being an arbitrary set, we easily see that $U^2 \mid R \mid U^2 = U^2$ for every relation $R \subseteq U^2$ such that $R \neq \Lambda$. Hence, by 4.10 and 4.23, every proper relation algebra $\mathfrak{A}'$ on U is simple, and $\mathfrak{A} \cong \mathfrak{A}'$ is a sufficient condition for $\mathfrak{A}$ to be simple. Assume now, conversely, that

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, \mid, I, {}^{-1} \rangle$$

is simple. By 4.24(i) we have

(1) V is a non-empty equivalence relation,

and in view of 4.27, we may assume that

(2) I is the identity function on the field of V.

Take a fixed element a in the field of V and put

(3) $$U = V^*(\{a\}).$$

Let B be the set of all relations $R \subseteq V$. It follows from (1), (2), and 4.26(i) that

$$\mathfrak{B} = \langle \mathsf{B}, \cup, \Lambda, \cap, V, \mid, I, {}^{-1} \rangle$$

is a proper relation algebra. (1) and (3) imply that

$$V \mid U^2 \mid V = U^2$$

so that, by 4.5(iv), U^2 is an ideal element in $\mathfrak{B}$. Hence, by 4.9 and 4.23, the function F on B defined by the formula

$$F(R) = U^2 \cap R \text{ for any } R \subseteq V$$

maps $\mathfrak{B}$ homomorphically onto the proper relation algebra

$$\mathfrak{B}(U^2) = \langle \mathsf{B}', \cup, \Lambda, \cap, U^2, \mid, I', {}^{-1} \rangle$$

where B' is the set of all relations $R \subseteq U^2$ while

(4) $$I' = U^2 \cap I$$

$\mathfrak{A}$ is clearly a subalgebra of $\mathfrak{B}$, and hence the same function F maps $\mathfrak{A}$ homomorphically onto the subalgebra

(5) $$\mathfrak{A}' = \langle \mathsf{A}', \cup, \Lambda, \cap, U^2, \mid, I', {}^{-1} \rangle$$

of $\mathfrak{B}(U^2)$, where $\mathsf{A}' = F^*(\mathsf{A})$. Moreover, by (1)-(4), the set U is not empty and I' is the identity function on U. Hence, according to 4.23, $\mathfrak{A}'$ is a proper relation algebra on U. Finally, since $\mathfrak{A}$ is simple and since, by (5), A' has at least two different elements ($U^2 \neq \Lambda$), F maps $\mathfrak{A}$ isomorphically onto $\mathfrak{A}'$. Thus the theorem holds in both directions.

THEOREM 4.29. *If*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

is an atomistic relation algebra in which every atom is a functional element, then $\mathfrak{A}$ is isomorphic to a proper relation algebra.

Proof. We repeat the proof of 4.22 by defining U, F, A, V, and I, and deriving conditions (6), (7), (8), (11), and (12). In view of 4.25, our task reduces to showing that

$$F(x \cdot y) = F(x) \cap F(y) \text{ for any } x, y \,\varepsilon\, A.$$

This condition is clearly equivalent to

(1) $\qquad (x \cdot y) ; b = (x ; b) \cdot (y ; b)$ for any $x, y \,\varepsilon\, A$ and $b \,\varepsilon\, U$.

Now, if $b \,\varepsilon\, U$, then $b^{\smile} \,\varepsilon\, U$ by 4.3(xii). Hence, by hypothesis, $b^{\smile}$ is a functional element, and we obtain (1) by applying 4.3(v)(vii) and 4.6(v).

THEOREM 4.30. *For every algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

the following two conditions are equivalent:

(i) *$\mathfrak{A}$ is isomorphic to a proper relation algebra*

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle$$

where A is the family of all relations $R \subseteq V$.

(ii) *$\mathfrak{A}$ is a complete, atomistic relation algebra in which every atom x satisfies the formula $x^{\smile} ; 1 ; x \leq 1'$.*

Conditions (i) *and* (ii) *remain equivalent if both the relation V in* (i) *is assumed to be of the form $V = U^2$ for some non-empty set U, and the algebra $\mathfrak{A}$ in* (ii) *is assumed to be simple.*[21]

Proof. Assume (i) to hold. Clearly, $\mathfrak{A}'$ is a complete atomistic relation algebra. Every atom $R \,\varepsilon\, \mathsf{A}$ obviously consists of a single ordered couple $\langle x, y \rangle \,\varepsilon\, V$; since, by 4.26(i), I is the identity function on the field of V, we easily check that $R^{-1} \,|\, V \,|\, R \subseteq I$. Hence (ii) follows at once.

---

[21] A different characterization of algebraic systems which are isomorphic to proper relation algebras constituted by all subrelations of a relation $V = U^2$ is given in McKinsey [2]. Theorem 4.30 implies that the algebraic systems discussed by McKinsey are relation algebras in the sense of 4.1, and in fact that they coincide with complete, atomistic, simple relation algebras in which every atom satisfies the formula $x^{\smile} ; 1 ; x \leq 1'$.

Assume now that (ii) holds. We want first to establish some properties of atoms in $\mathfrak{A}$. Let U be the set of all atoms in $\mathfrak{A}$. From (ii) and 4.3(ii)(iv)(vii)(xii) we easily obtain:

(1) if $u \,\varepsilon\, U$, then $u^{\smile} \,\varepsilon\, U$, $u^{\smile} ; u \leq 1'$, and $u ; u^{\smile} \leq 1'$.

Hence, by 4.5(iii) and 4.19,

(2) if $u \,\varepsilon\, U$, then $u^{\smile} ; u \,\varepsilon\, U$ and $u ; u^{\smile} \,\varepsilon\, U$,

and

(3) if $v, w \,\varepsilon\, U$, then $v ; w \,\varepsilon\, U$ or $v ; w = 0$.

Consider any $u, v \,\varepsilon\, U$ such that $u ; u^{\smile} = v ; v^{\smile}$ and $u^{\smile} ; u = v^{\smile} ; v$. We then clearly have

$$u ; u^{\smile} ; v = v ; v^{\smile} ; v = v ; u^{\smile} ; u.$$

Hence, by 4.3(iii)(ix), $u^{\smile} ; v \neq 0$ and $v ; u^{\smile} \neq 0$. Thus $(v ; u^{\smile}) \cdot 1 \neq 0$ whence, by 4.1(iv), $(v^{\smile} ; 1) \cdot u^{\smile} \neq 0$. Therefore, by (1), 4.3(ii)(iv), and (ii), $u^{\smile} \leq v^{\smile} ; 1$ and $u^{\smile} ; v \leq v^{\smile} ; 1 ; v \leq 1'$. Since $u^{\smile} ; v \neq 0$, we obtain $(u^{\smile} ; v) \cdot 1' \neq 0$. By applying 4.1(iii)(iv) we conclude that $(u ; 1') \cdot v \neq 0$ and $u \cdot v \neq 0$, so that finally $u = v$. Thus,

(4) if $u, v \,\varepsilon\, U$, $u ; u^{\smile} = v ; v^{\smile}$, and $u^{\smile} ; u = v^{\smile} ; v$, then $u = v$.

Now consider any $u, v, w \,\varepsilon\, U$ such that $u \leq v ; w$. By 4.1(iii), 4.3(ii) (v) (vi), and (1), we conclude that

$$u ; u^{\smile} \leq v ; w ; (v ; w)^{\smile} \leq v ; w ; w^{\smile} ; v^{\smile} \leq u ; 1' ; v^{\smile} \leq v ; v^{\smile}.$$

Hence, by (2), $u ; u^{\smile} = v ; v^{\smile}$. Similarly we obtain $u^{\smile} ; u = w^{\smile} ; w$. Since $u \neq 0$, we have

$$v ; w = (v ; w) \cdot (v ; w) \neq 0.$$

By applying 4.1(iv) twice we arrive at

$$(v^{\smile} ; v ; w) \cdot w \neq 0 \text{ and } (w ; w^{\smile}) \cdot (v^{\smile} ; v) \neq 0.$$

Hence. by (2), $w ; w^{\smile} = v^{\smile} ; v$. Thus,

(5) if $u, v, w \,\varepsilon\, U$ and $u \leq v ; w$, then $u ; u^{\smile} = v ; v^{\smile}$, $u^{\smile} ; u = w^{\smile} ; w$, and $v^{\smile} ; v = w ; w^{\smile}$.

Finally consider any $v, w \,\varepsilon\, U$ such that $v^{\smile} ; v = w ; w^{\smile}$, and let $u = v ; w$. Since $(v^{\smile} ; v) \cdot (w ; w^{\smile}) \neq 0$, we obtain, by 4.1(iv), $(v ; w ; w^{\smile}) \cdot v \neq 0$. Hence, by 4.3(iii) and (3), $u = v ; w \neq 0$ and $u \,\varepsilon\, U$. Thus, in view of (5),

(6) if $v, w \,\varepsilon\, U$, $v^{\smile} ; v = w ; w^{\smile}$, and $u = v ; w$, then $u \,\varepsilon\, U$, $u ; u^{\smile} = v ; v^{\smile}$, and $u^{\smile} ; u = w^{\smile} ; w$.

We now define a function F on A by putting for every $x \,\varepsilon\, A$.

(7) $$F(x) = \mathop{\mathrm{E}}_{\langle a,b\rangle} [\textit{for some } u \,\varepsilon\, U, \quad u \leqq x, \quad a = u ; u^{\smile}, \textit{ and } b = u^{\smile} ; u].$$

We also put

(8) $$F^*(A) = \mathsf{A}, \quad F(1) = V, \quad \text{and} \quad F(1') = I.$$

If $x, y \,\varepsilon\, A$ and $F(x) = F(y)$, we conclude from (4) and (7) that the formulas $u \leq x$ and $u \leq y$ are equivalent for every $u \,\varepsilon\, U$. Hence, $\mathfrak{A}$ being atomistic,

(9) $$F(x) = F(y) \text{ implies } x = y \text{ for any } x, y \,\varepsilon\, A.$$

Furthermore, we obtain from (7) with the help of (4),

(10) $$F(x + y) = F(x) \cup F(y) \text{ and } F(x \cdot y) = F(x) \cap F(y)$$ for any $x, y \,\varepsilon\, A$,

(11) $$F(0) = \Lambda.$$

Also, by 4. 3 (v) (vi) (vii) and (2), we conclude from (7) that

(12) $$F(x^{\smile}) = [F(x)]^{-1} \text{ for every } x \,\varepsilon\, A.$$

Let now $x, y \,\varepsilon\, A$, and assume $\langle a, b\rangle \,\varepsilon\, F(x ; y)$. By (7) we have $u \leq x ; y$, $a = u ; u^{\smile}$, and $b = u^{\smile} ; u$ for some $u \,\varepsilon\, U$. Hence, by (ii) and with the help of 4. 3 (i), $u \leq v ; w$ for some $v, w \,\varepsilon\, U$ such that $v \leq x$ and $w \leq y$. By (5) we obtain $a = v ; v^{\smile}$, $b = w^{\smile} ; w$, and $v^{\smile} ; v = w ; w^{\smile}$. Let $c = v^{\smile} ; v = w ; w^{\smile}$. Then, by (7), $\langle a, c\rangle \,\varepsilon\, F(x)$, $\langle c, b\rangle \,\varepsilon\, F(y)$, and consequently $\langle a, b\rangle \,\varepsilon\, F(x) \mid F(y)$. If, conversely, $\langle a, b\rangle \,\varepsilon\, F(x) \mid F(y)$, then by means of an analogous argument, but going in the opposite direction and using (6) instead of (5), we obtain $\langle a, b\rangle \,\varepsilon\, F(x ; y)$. Hence

(13) $$F(x ; y) = F(x) \mid F(y) \text{ for any } x, y \,\varepsilon\, A.$$

From (8)-(13) we see that the function F maps isomorphically the relation algebra $\mathfrak{A}$ onto the algebra

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, V, \mid, I, {}^{-1}\rangle$$

Hence, by 4. 1 and 4. 23, $\mathfrak{A}'$ is a proper relation algebra. The elements of A are relations $R \subseteq V$. Consider any relation $R \subseteq V$. By (7) and (8), for every $\langle a, b\rangle \,\varepsilon\, R$ there is a $u \,\varepsilon\, U$ such that $a = u ; u^{\smile}$ and $b = u^{\smile} ; u$. Since, by (ii), $\mathfrak{A}$ is complete, the sum X of all $u \,\varepsilon\, U$ thus obtained exists, and we easily see from (7) that $F(x) = R$; hence, $R \varepsilon \mathsf{A}$. Therefore A consists of all relations $R \subseteq V$, and $\mathfrak{A}'$ is an algebra of the kind described in (i).

Thus conditions (i) and (ii) are equivalent. By 4.26(i)(ii), these conditions remain equivalent if they are modified in the way indicated in the last part of the theorem.

THEOREM 4.31. *For every relation algebra $\mathfrak{B}$ the following conditions are equivalent:*

(i) *$\mathfrak{B}$ is isomorphic to a proper relation algebra $\mathfrak{B}'$.*

(ii) *$\mathfrak{B}$ is a subalgebra of a complete atomistic relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

in which every atom x satisfies the formula $x^{\smile} ; 1 ; x \leq 1'$.

(iii) *$\mathfrak{B}$ is a subalgebra of an atomistic relation algebra $\mathfrak{A}$ in which every atom is a functional element.*

Conditions (i)-(iii) *remain equivalent if both the algebra $\mathfrak{B}'$ in* (i) *is assumed to be a proper relation algebra on a set U, and the algebra $\mathfrak{A}$ in* (ii) *and* (iii) *is assumed to be simple.*

Proof. Assume (i) to hold, and put, according to 4.23,

$$\mathfrak{B}' = \langle \mathsf{B}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle.$$

By 4.24(i), V is an equivalence relation. By 4.27, I may be assumed to be the identity function on the field of V. Let A be the family of all relations $R \subseteq V$. Then, by 4.26,

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle$$

is a proper relation algebra, and $\mathfrak{B}'$ is clearly a subalgebra of $\mathfrak{A}'$. Since $\mathfrak{B} \cong \mathfrak{B}'$, we can construct, by applying the familiar "exchange method," an algebra $\mathfrak{A} \cong \mathfrak{A}'$ such that $\mathfrak{B}$ is a subalgebra of $\mathfrak{A}$. By 4.30, $\mathfrak{A}$ is a complete atomistic relation algebra in which every atom x satisfies the formula $x^{\smile} ; 1 ; x \leq 1'$. Thus, condition (i) implies (ii). By 4.3(ii)(iv) and 4.5(iii), condition (ii) clearly implies (iii), while, by 4.29, (iii) implies (i). Hence, conditions (i)-(iii) are equivalent. By means of an analogous argument we can show that these conditions remain equivalent if they are modified in the way indicated in the last part of the theorem; to derive then (i) from (iii), we apply 4.11 and 4.28, in addition to 4.29.

So far Theorem 4.15 has not been involved in our discussion of the representation problem. Nevertheless, some possibilities of applying this theorem to the representation problem can easily be anticipated. In fact, as a consequence of 4.15, the representation problem for arbitrary relation algebras reduces to that for simple relation algebras. Speaking more specifically, consider any class K of relation algebras, and let L be the class of all simple relation algebras which are homomorphic images of algebras in K. It is easily seen that, if every relation algebra of a certain class has a natural representation, the same applies to all cardinal products and to all subalgebras of algebras of this class. Hence, if we succeed in showing that every algebra of the class L has a natural representation, then, due to 4.15, this result automatically extends to all algebras of the class K. We do not know, however, any interesting applications of Theorem 4.15 in the direction just indicated. On the other hand, it will be seen from our further discussion that in some cases by means of Theorem 4.15 we can obtain additional information regarding relation algebras of which we have been able to show (without the help of this theorem) that they have a natural representation.

THEOREM 4.32. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

be a relation algebra in which the element 1 *can be represented in the form*

$$1 = \sum_{k<m} x_k$$

where $x_0, x_1, \cdots, x_{m-1}$ *are functional elements. Then:*

(i) $\mathfrak{A}$ *is isomorphic to a proper relation algebra.*

(ii) *If* $\mathfrak{A}$ *is simple, it is isomorphic to a proper relation algebra* $\mathfrak{A}'$ *on a set* U *which contains at most* m *different elements.*

(iii) *In the general case,* $\mathfrak{A}$ *is isomorphic to a subalgebra of a cardinal product of proper relation algebras* $\mathfrak{A}_i$ *on sets* U_i*, each of these sets containing at most* m *different elements.*

Proof. By 4.21, there is a complete and atomistic relation algebra

$$\mathfrak{B} = \langle B, +, 0, \cdot, ;, 1', \smile \rangle$$

which is a perfect extension of $\mathfrak{A}$. $\mathfrak{A}$ is of course a subalgebra of $\mathfrak{B}$ (cf. 2.14). The formula

$$1 = \sum_{i<m} x_i$$

stated in the hypothesis continues to hold in $\mathfrak{B}$ (since the summation is finite); and, by 4.5(iii), the elements x_i which are functional in $\mathfrak{A}$ are functional in $\mathfrak{B}$ as well. Therefore, for every atom $a \varepsilon B$ there is a functional element $x_i \varepsilon B$ such that $a \leq x_i$. Hence, by 4. 6(iv), every atom in $\mathfrak{B}$ is a functional element. Consequently, by 4. 31(i) (iii), the relation algebra $\mathfrak{A}$ is isomorphic to a proper relation algebra, and conclusion (i) has thus been established.

If, in addition, $\mathfrak{A}$ is simple, we see, by (i) and 4. 28, that $\mathfrak{A}$ is isomorphic to a proper relation algebra on a set U. From the hypothesis we conclude, by 4. 23 and 4. 24(ii), that U^2 can be represented as the union of m many-to-one relations. Hence, for every $x \varepsilon U$, there are at most m different elements y such that $\langle x, y\rangle \varepsilon U^2$. Consequently, U has at most m different elements, and conclusion (ii) is seen to hold.

In the general case (i. e., without assuming that $\mathfrak{A}$ is simple), we easily see by 4. 5(iii) that, not only the algebra $\mathfrak{A}$ itself, but also every homomorphic image of $\mathfrak{A}$ satisfies the hypothesis of our theorem. Hence, by applying 4. 15 and (ii), we obtain conclusion (iii).

THEOREM 4. 33. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile\rangle$$

be a relation algebra in which $0' ; 0' \leq 1'$. *Then:*

(i) *$\mathfrak{A}$ is isomorphic to a proper relation algebra.*

(ii) *If $\mathfrak{A}$ is simple, it is isomorphic to a proper relation algebra $\mathfrak{A}'$ on a set U which contains at most two different elements.*

(iii) *In the general case, $\mathfrak{A}$ is isomorphic to a subalgebra of a direct product of proper relation algebras $\mathfrak{A}_i$ on sets U_i, each of these sets containing at most two different elements.*

Proof. By 4. 5(i) we obviously have $1 = 1' + 0'$. By 4. 6(vi), the element $1'$ is functional; under the hypothesis of our theorem, in view of 4. 5(iii) and 4. 6(i), the element $0'$ is also functional. Thus, 1 is a sum of two functional elements. By now applying 4. 32, we obtain all the conclusions at once.

To formulate conveniently the last two theorems of this section, we introduce the following notation:

DEFINITION 4. 34. *A relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile\rangle$$

is said to be

(i) *of class 1, if* $0' ; 0' = 0$,

(ii) *of class 2, if* $0' ; 0' = 1'$,

(iii) *of class 3, if* $0' ; 0' = 1$.

THEOREM 4. 35. *Every simple relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

satisfies one and only one of the following three conditions:

(i) $\mathfrak{A}$ *is of class* 1 *and is isomorphic to a proper relation algebra* $\mathfrak{A}'$ *on a set* U *which contains just one element*;

(ii) $\mathfrak{A}$ *is of class* 2 *and is isomorphic to a proper relation algebra* $\mathfrak{A}'$ *on a set* U *which contains just two different elements*;

(iii) $\mathfrak{A}$ *is of class* 3 *and either is not isomorphic to any proper relation algebra or else is isomorphic to a proper relation algebra* $\mathfrak{A}'$ *on a set* U *which contains at least three different elements.*

Proof. From 4. 5(i), 4. 23 and 4. 34(i)(ii) we easily conclude:

(1) Every proper relation algebra $\mathfrak{A}'$ on a set U containing just one element, or just two different elements, is of class 1, or of class 2, respectively.

Hence, by 4. 33 and 4. 34, if the given relation algebra $\mathfrak{A}$ satisfies the formula $0' ; 0' \leq 1'$, then it satisfies one of the conditions (i) or (ii).

If $\mathfrak{A}$ does not satisfy the above formula, we have, by 4. 5(i), $(0' ; 0') \cdot 0' \neq 0$ and therefore, by 4. 10,

$$1 ; ((0' ; 0') \cdot 0') ; 1 = 1.$$

Hence we obtain, by 4. 1(iii), 4. 3(i)(ii), 4. 5(i), and 4. 6(ii),

$$\begin{aligned} 1 &= (0' + 1') ; ((0' ; 0') \cdot 0') ; (0' + 1') \\ &\leq 0' ; 0' ; 0' ; 0' + 0' ; 0' + 0' ; 0' + 0' ; 0' = 0' ; 0'. \end{aligned}$$

Thus, by 4. 34(iii), $\mathfrak{A}$ is in this case of class 3. Also, since the formula $0' ; 0' \leq 1'$ fails, it follows from (1) and 4. 34 that $\mathfrak{A}$ cannot be isomorphic to any proper relation algebra on a set U containing at most two different elements. Hence, $\mathfrak{A}$ being simple by hypothesis, we conclude with the help of 4. 28 that $\mathfrak{A}$ satisfies condition (iii). This completes the proof.

From Theorem 4. 35 it is easily seen that, up to isomorphism, there is only one simple relation algebra of class 1 and there are only two simple relation algebras of class 2. One of these two simple algebras of class 2 has exactly four elements (e. g., the relations Λ, I, D, and U^2, where U is a set containing just two elements, I is the identity function on U, and D is the complement of I to U^2); the other has exactly sixteen elements (e. g., all the relations $R \subseteq U^2$ where U is a set containing just two elements).

As regards arbitrary relation algebras, it easily follows from 4. 34 that none of them can be both of class m and class n for $1 \leqq m < n \leqq 3$ unless it is a one-element algebra. On the other hand, there are relation algebras which are not of class n for any $n = 1, 2, 3$. By 4.15 and 4.34, a relation algebra is of class n if, and only if, it is a subalgebra of a cardinal product of simple relation algebras of class n. In particular, relation algebras of class 1 are of a rather trivial nature. They can be characterized by the simple formula $1' = 1$, or else by the condition that every element in such an algebra is an ideal element. It is seen from 4.1 that every Boolean algebra $\langle B, +, 0, \cdot, 1\rangle$ can be completed to a relation algebra $\langle B, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$ of class 1 by putting $x ; y = x \cdot y$ and $x^{\smile} = x$ for any $x, y \varepsilon A$, as well as $1' = 1$. For this reason relation algebras of class 1 are also referred to as *Boolean relation algebras.*

Since a cardinal product of relation algebras of class n $(n = 1, 2, 3)$ is again a relation algebra of class n, we infer from 4.15 and 4.35 that every relation algebra $\mathfrak{A}$ can be represented as a subalgebra of the cardinal product of three algebras $\mathfrak{A}_1$, $\mathfrak{A}_2$, and $\mathfrak{A}_3$ of classes 1, 2, and 3, respectively. This result, however, will be essentially improved in the following

THEOREM 4.36. *Every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile}\rangle$$

is isomorphic to the cardinal product of relation algebras $\mathfrak{A}_1$, $\mathfrak{A}_2$, *and* $\mathfrak{A}_3$ *of classes* 1, 2, *and* 3 *respectively.*

Proof. We put

(1) $$a_3 = 1 ;((0' ; 0') \cdot 0') ; 1, \qquad a_2 = (1 ; 0' ; 1) \cdot a_3^-,$$
$$\text{and } a_1 = a_2^- \cdot a_3^-.$$

Hence obviously

(2) $$1 = a_1 + a_2 + a_3 \quad \text{and} \quad a_1 \cdot a_2 = a_1 \cdot a_3 = a_2 \cdot a_3 = 0.$$

Furthermore, by (1) and 4.6(viii) (xii),

(3) a_1, a_2, and a_3 are ideal elements.

By 4.9 and 4.12, conditions (2) and (3) imply that $\mathfrak{A}(a_1)$, $\mathfrak{A}(a_2)$, and $\mathfrak{A}(a_3)$ are relation algebras such that

(4) $$\mathfrak{A} \cong \mathfrak{A}(a_1) \times \mathfrak{A}(a_2) \times \mathfrak{A}(a_3).$$

From 4.8 or 4.9 we see that $0, a_i, a_i \cdot 1'$, and $a_i \cdot 0'$ are respectively the zero element, the unit element, the identity element, and the diversity element of the algebras $\mathfrak{A}(a_i)$, $i = 1, 2, 3$. Hence, in view of (4) and 4.34, in order to complete the proof it suffices to show that

(5) $$(a_1 \cdot 0') \mathbin{;} (a_1 \cdot 0') = 0,$$

(6) $$(a_2 \cdot 0') \mathbin{;} (a_2 \cdot 0') = a_2 \cdot 1',$$

and

(7) $$(a_3 \cdot 0') \mathbin{;} (a_3 \cdot 0') = a_3.$$

These three formulas can be established directly by using the laws of the arithmetic of relation algebras stated in 4.1, 4.3, and 4.6. The following indirect proof is, however, somewhat simpler and shorter.

Assume first that $\mathfrak{A}$ is a simple relation algebra. Then, by 4.35, $\mathfrak{A}$ is of class 1, 2, or 3. If $\mathfrak{A}$ is of class 1, we have, by 4.3(iii)(ix)(xi), 4.5(i), and 4.34,

$$0' \leqq 0' ; 0' ; 0' = 0 ; 0' = 0$$

so that $0' = 0$. Hence, by (1), 4.3(iii), and 4.34, $a_3 = 0$, $a_2 = 0$, and $a_1 = 1$. If $\mathfrak{A}$ is of class 2, then it follows from (1), 4.3(iii), 4.10(i)(iii), and 4.34 that $a_3 = 0$, $a_2 = 1$, and $a_1 = 0$. Finally, in case $\mathfrak{A}$ is of class 3, we infer from (1) and 4.10(i)(iii) that $a_3 = 1$, $a_2 = 0$, and $a_1 = 0$. In each of these three cases it easily follows from 4.3(iii) and 4.34 that (5)-(7) are satisfied.

It is easily seen that whenever formulas (5)-(7) hold in given relation algebras, they also hold in cardinal products and subalgebras of these algebras. (No essential difficulty arises from the fact that, in view of (1) and 4.5(i), these formulas implicitly involve the operation of complementation, which is not included in the system of fundamental operations of relation algebras.) Consequently, since formula (5)-(7) have been shown to hold in simple relation algebras, we conclude by means of 4.15 that they also hold in arbitrary relation algebras, and the proof is complete.

To illustrate this theorem, consider the case when $\mathfrak{A}$ is a proper relation algebra,

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, V, |, I, {}^{-1} \rangle$$

where V is an equivalence relation with the field U, and I is the identity function on U. Let K be the partition of U under V (see remarks following 4.24). Let K_1, K_2, and K_3 be respectively the families of all those sets $X \in \mathsf{K}$ which have exactly one, exactly two, and at least three different elements. Finally, let

$$V_i = \bigcup_{X \in \mathsf{K}_i} X^2 \text{ for } i = 1, 2, 3.$$

The relations V_i prove to be elements of A; in fact, they can be identified with the elements a_i defined in the proof of 4. 35. It is easily seen that they are ideal elements of $\mathfrak{A}$; they are obviously disjoint, and their union is V. Hence

$$\mathfrak{A} \cong \mathfrak{A}(V_1) \times \mathfrak{A}(V_2) \times \mathfrak{A}(V_3).$$

Finally, by means of a simple direct argument, we can show that the algebra $\mathfrak{A}(V_i)$, for $i = 1, 2, 3$, is of the class i in the sense of 4. 34.

From 4. 36 we can derive some further consequences by means of 4. 15 and 4. 34. For instance, it is easily seen that the formula $0' ; 0' \leq 1'$ characterizes those relation algebras $\mathfrak{A}$ for which the algebra $\mathfrak{A}_3$ of 4. 36 has just one element, and which therefore can be represented as cardinal products of an algebra A_1 of class 1 and an algebra $\mathfrak{A}_2$ of class 2; such algebras can also be characterized (up to isomorphism) as subalgebras of cardinal products of simple relation algebras of classes 1 and 2. Similarly, the formula $0' \leq 0' ; 0'$ characterizes those relation algebras which are isomorphic to cardinal products of an algebra $\mathfrak{A}_1$ of class 1 and an algebra $\mathfrak{A}_3$ of class 3. Finally, the formula $1' \leq 0' ; 0'$ is characteristic for those relation algebras which are representable as cardinal products of an algebra $\mathfrak{A}_2$ of class 2 and an algebra $\mathfrak{A}_3$ of class 3.

Section 5.

Relation Algebras and Brandt Groupoids.

In Section 3 we have established fundamental relations between Boolean algebras with operators and complex algebras of arbitrary algebraic systems (algebras in the wider sense). In Section 4 we have studied a special class of Boolean algebras with operators—in fact, the relation algebras. We now want to discuss connections between relation algebras and a special class of algebraic systems—in fact the (generalized) Brandt groupoids; this class includes in particular all the groups as its members.

A Brandt groupoid is an algebraic system formed by a set U of elements, a binary operation $\cdot$, a distinguished subset of U (the set of identity elements), and a unary operation $^{-1}$. Roughly speaking, the main difference between Brandt groupoids and groups consists in the fact that in a Brandt groupoid the set U is not assumed to be closed under multiplication; the domain of the operation $\cdot$ is not, in general, the whole set U^2, but a subset of U^2. In consequence, a Brandt groupoid may contain many identity elements, i. e., many elements u such that $x \cdot u = x = u \cdot x$ whenever x, $x \cdot u$, and $u \cdot x$ are in U. A precise definition of a Brandt groupoid follows:

DEFINITION 5.1. *An algebraic system (algebra in a wider sense)*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

(where $\cdot$ is an operation on a subset of U^2 to U, I is a subset of U, and ${}^{-1}$ is an operation on U to U) is called a generalized Brandt groupoid if the following conditions are satisfied:

(i) *For any elements $x, y, z \in U$ such that $x \cdot y \in U$ and $y \cdot z \in U$ we have $(x \cdot y) \cdot z \in U$ and $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.*

(ii) *For any elements $x, y, z \in U$ such that $x \cdot y \in U$ and $x \cdot y = x \cdot z$ we have $y = z$.*

(iii) *For any elements, $x, y, z \in U$ such that $x \cdot z \in U$ and $x \cdot z = y \cdot z$ we have $x = y$.*

(iv) *$x \cdot x = x$ for every $x \in I$.*

(v) *$x^{-1} \cdot x \in I$ and $x \cdot x^{-1} \in I$ for every $x \in U$.*

U is called a Brandt groupoid if, in addition, the following condition holds:

(vi) *For any elements $x, z \in I$ there exists an element $y \in U$ such that $x \cdot y \in U$ and $y \cdot z \in U$.*[22]

In the next three theorems we state without proof certain arithmetic properties of generalized Brandt groupoids which will be used later. These results can be obtained by methods that are essentially known from the literature.[23]

THEOREM 5.2. *Let*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

be a generalized Brandt groupoid. For any elements $x, y, z \in U$ we have:

(i) *If $y \in I$ and $x \cdot y \in U$, then $x \cdot y = x$.*

(ii) *If $x \in I$ and $x \cdot y \in U$, then $x \cdot y = y$.*

(iii) *$x \cdot (x^{-1} \cdot x) = x = (x \cdot x^{-1}) \cdot x$.*

(iv) *If $x \cdot y \in I$, then $x^{-1} = y$.*

(v) *$(x^{-1})^{-1} = x$.*

(vi) *If $x \cdot y \in U$, then $x^{-1} \cdot x = y \cdot y^{-1}$.*

(vii) *If $x \cdot y \in U$, then $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$.*

---

[22] These algebraic systems were first studied by H. Brandt. It is easily seen that axioms (i)-(vi) in Definition 5.1 are equivalent to axioms I-IV of Brandt [1]; when deriving 5.1(i)-(vi) from Brandt's axioms, we let I be the set of all elements $x \in U$ such that $x \cdot x = x$.

[23] Cf. Brandt [1].

THEOREM 5.2. *An algebraic system*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

(where $\cdot$ is a function on a subset of U^2 to U, $I \subseteq U$, and ${}^{-1}$ is a function on U to U) is a generalized Brandt groupoid if, and only if, the following conditions are satisfied:

(i) *For every element $x \varepsilon U$ there is an element $y \varepsilon U$ such that $x \cdot y \varepsilon U$.*

(ii) *For any elements $x, y, z \varepsilon U$, if $x \cdot y$ and $(x \cdot y) \cdot z$ are in U or $y \cdot z$ and $x \cdot (y \cdot z)$ are in U, then all these elements are in U and $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.*

(iii) *For any element $x \varepsilon U$ we have $x \varepsilon I$ if, and only if, $x \cdot x = x$.*

(iv) *For any elements $x, y, z \varepsilon U$ the formulas*

$$x \cdot y = z, \quad x^{-1} \cdot z = y, \text{ and } z \cdot y^{-1} = x$$

are equivalent.

THEOREM 5.4. *For every generalized Brandt groupoid*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

the following three conditions are equivalent:

(i) *$\mathfrak{U}$ is a Brandt groupoid.*

(ii) *For any elements $x, z \varepsilon U$ there is an element $y \varepsilon U$ such that $x \cdot y \varepsilon U$ and $y \cdot z \varepsilon U$.*

(iii) *For any elements $x, y \varepsilon U$ there are elements $u, v \varepsilon U$ such that $u \cdot y \varepsilon U$ and $x = u \cdot y \cdot v$.*

According to Definition 3.8, the complex algebra of a generalized Brandt groupoid

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

is the algebra

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, \cdot^*, I^*, {}^{-1*} \rangle$$

where A is the family of all subsets of U. The operations $\cdot^*$, I^*, and ${}^{-1*}$ are understood in the sense of Definition 3.2. Thus, $\cdot^*$ is an operation on A^2 to A; for any sets X and Y in A, $X \cdot^* Y$ is the set of all elements $z \varepsilon U$ such that $z = x \cdot y$ for some $x \varepsilon X$ and $y \varepsilon Y$. Similarly, for any $X \varepsilon \mathsf{A}$, X^{-1*} is the set of all $y \varepsilon U$ such that $y = X^{-1}$ for some $x \varepsilon X$. Finally, since I is a subset of U, i. e., a unary relation, I^* is an operation of rank 0 such that $I^*(\Lambda) = I$. This operation I^* will be replaced as usual by the set I

l.c.

itself which will be treated as a distinguished element in A (see remarks in the introduction). Thus we shall speak of

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, \cdot^*, I, {}^{-1*} \rangle$$

as the complex algebra of $\mathfrak{U}$.

THEOREM 5.5. *Let*

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

be a generalized Brandt groupoid, and let

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, \cdot^*, I, {}^{-1*} \rangle$$

be the complex algebra of $\mathfrak{U}$. *Then we have:*

(i) $\mathfrak{A}$ *is a relation algebra with at least two different elements* ($\Lambda \neq U$)—*in fact, a complete atomistic relation algebra in which every atom is a functional element.*

(ii) *If* $\mathfrak{U}$ *is a Brandt groupoid, then* $\mathfrak{A}$ *is simple.*

Proof. By 3.8, $\mathfrak{A}$ is a complete atomistic Boolean algebra with operators. By 5.3(ii), for any sets $X, Y, Z \,\varepsilon\, \mathsf{A}$ (i.e., for any subsets X, Y, Z of U) we have

$$(X \cdot^* Y) \cdot^* Z = X \cdot^* (Y \cdot^* Z).$$

From 5.1(v) and 5.2(i)-(iii) we conclude that

$$I \cdot^* X = X = X \cdot^* I.$$

By 5.3(iv) the formulas

$$(X \cdot^* Y) \cap Z = \Lambda, \ (X^{-1*} \cdot^* Z) \cap Y = \Lambda, \text{ and } (Z \cdot^* Y^{-1*}) \cap X = \Lambda$$

are equivalent. Hence, by 4.1, $\mathfrak{A}$ is a relation algebra. Since $U \neq \Lambda$, $\mathfrak{A}$ has at least two different elements.

Every atom X in $\mathfrak{A}$ is clearly a set of the form $X = \{x\}$ for some $x \,\varepsilon\, U$, and therefore, by 5.1(v), it satisfies the formula $X^{-1*} \cdot^* X \subseteq I$. Thus, by 4.5(iii), every atom in $\mathfrak{A}$ is a functional element.

If, finally, $\mathfrak{U}$ is a Brandt groupoid, we easily see from 5.4(i) (iii) that $U \cdot^* X \cdot^* U = U$ for every $X \,\varepsilon\, \mathsf{A}$ such that $X \neq \Lambda$. Hence, by 4.10(i)(iii), $\mathfrak{A}$ is simple.

THEOREM 5.6. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', {}^{\smile} \rangle$$

be an atomistic relation algebra, with $0 \neq 1$, *in which every atom is a func-*

tional element. Let U be the set of all atoms of $\mathfrak{A}$, let I be the set of all those atoms u for which $u \leqq 1'$, and let

$$\mathfrak{U} = \langle U, ;, I, {}^\smile \rangle.$$

We then have:

(i) *$\mathfrak{U}$ is a generalized Brandt groupoid, and $\mathfrak{A}$ is isomorphic to a subalgebra of the complex algebra of $\mathfrak{U}$.*

(ii) *If $\mathfrak{A}$ is complete, then $\mathfrak{A}$ is isomorphic to the complex algebra of $\mathfrak{U}$.*

(iii) *If $\mathfrak{A}$ is simple, then $\mathfrak{U}$ is a Brandt groupoid.*

Proof. We first want to show that the algebraic system $\mathfrak{U}$ satisfies condition 5.3(i)-(iv), with $\cdot$ and $^{-1}$ replaced by $;$ and ${}^\smile$, respectively. (The domain of the operation $;$ is understood to be restricted to the set of those couples $\langle x, y \rangle \varepsilon U^2$ for which $x ; y \varepsilon U$. An analogous remark applies to the operation ${}^\smile$, and from 4.3(xii) it is seen that the domain of the operation ${}^\smile$ thus restricted is U.)

By 4.3(vii)(xii) and 4.19(i)(iii) we have $x^\smile \varepsilon U$ and $x ; x^\smile \varepsilon U$ for every $x \varepsilon U$. Hence 5.3(i) holds. If, x, y, z, $x ; y$, and $(x ; y) ; z$ are in U, we have $(x ; y) ; z = x ; (y ; z)$ by 4.1(ii). Hence, by 4.3(iii), we conclude that $y ; z \neq 0$, and therefore $y ; z \varepsilon U$ by 4.19(i) (ii). Similarly, under the assumption that x, y, z, $y ; z$, and $x ; (y ; z)$ are in U, we obtain $(x ; y) ; z = x ; (y ; z)$ and $x ; y \varepsilon U$. Thus, 5.3(ii) holds. From 4.6(iii)(vi) we see that $x ; x = x$ for every $x \varepsilon I$. If, conversely, $x ; x = x$ and $x \varepsilon U$, we have, by 4.1(iii), 4.6(v), and the hypothesis of the theorem, $x = (x ; x) \cdot (x ; 1') = x ; (x \cdot 1')$. Hence, by 4.3(iii), $x \cdot 1' \neq 0$, therefore $x \cdot 1' = x$ (x being an atom), and consequently $x \varepsilon I$. We have thus obtained 5.3(iii). If, finally, $x, y, z \varepsilon U$ and $x ; y = z$, then $(x ; y) \cdot z \neq 0$, hence, by 4.1(iv), $(x^\smile ; z) \cdot y \neq 0$, and therefore $x^\smile ; z \neq 0$. Consequently, by 4.3(xii) and 4.19, $x^\smile ; z$ is an atom, and since y is also an atom, the formula $(x^\smile ; z) \cdot y \neq 0$ gives $x^\smile ; z = y$. Thus $x ; y = z$ implies $x^\smile ; z = y$ for any $x, y, z \varepsilon U$. From this implication we easily derive 5.3(iv) by means of 4.3(v)(vii).

We now know by 5.3 that $\mathfrak{U}$ is a generalized Brandt groupoid. We define a function F on A by putting for every $x \varepsilon A$

(1) $$F(x) = \underset{u}{\mathrm{E}} \, [u \varepsilon U \text{ and } u \leq x].$$

Since $\mathfrak{A}$ is atomistic, this definition clearly implies for any $x, y \varepsilon A$:

(2) If $F(x) = F(y)$, then $x = y$ (i.e., F is one-to-one).

(3) $F(x + y) = F(x) \cup F(y)$ and $F(x \cdot y) = F(x) \cap F(y)$.

We also have

(4) $$F(0) = \Lambda,\quad F(1) = U, \text{ and } F(1') = I.$$

By 4. 3(vi)(vii)(xii) we obtain for every $x \varepsilon A$

(5) $$F(x^{\smile}) = \underset{v}{\mathrm{E}}\,[v \varepsilon U \textit{ and, for some } u,\quad v = u^{\smile} \textit{ and } u \varepsilon F(x)].$$

Finally, we can show that for any $x, y \varepsilon A$

(6) $$F(x\,;y) = \underset{w}{\mathrm{E}}\,[w \varepsilon U \textit{ and, for some } u \textit{ and } v,\quad w = u\,;v,\quad u \varepsilon F(x), \textit{ and } v \varepsilon F(y)].$$

In fact, if $w \varepsilon F(x\,;y)$, then, by (1), $w \varepsilon U$ and $w \leq x\,;y$. Hence, with the help of 4. 3(i) and 4. 19, $w = u\,;v$ for some $u, v \varepsilon U$ such that $u \leq x$ and $v \leq y$, and therefore, again by (1), $w = u\,;v$ for some $u \varepsilon F(x)$ and $v \varepsilon F(y)$. If, conversely, $w \varepsilon U$ and $w = u\,;v$ for some $u \varepsilon F(x)$ and $v \varepsilon F(y)$, then $w \varepsilon F(x\,;y)$ by (1) and 4. 3(ii).

In view of 3. 2 and 3. 8, conditions (1)-(6) show that F maps $\mathfrak{A}$ isomorphically onto a subalgebra $\mathfrak{A}'$ of the complex algebra of $\mathfrak{U}$. If $\mathfrak{A}$ is complete, $\mathfrak{A}'$ clearly coincides with the complex algebra of $\mathfrak{U}$. If $\mathfrak{A}$ is simple, we have, by 4. 10, $x \leq 1\,;y\,;1$ for any $x, y \varepsilon U$, and hence, by 4. 3(i)(iii) and 4. 19, $x = u\,;y\,;v$ and $u\,;y \varepsilon U$ for some $u, v \varepsilon U$. By 5. 4 this implies that $\mathfrak{U}$ is a Brandt groupoid. The proof has thus been completed.

We now can establish a converse of 5. 5:

THEOREM 5. 7. *Let*

$$\mathfrak{U} = \langle U, R, I, S\rangle$$

be an algebra in the wider sense in which $R \subseteq U^3$, $I \subseteq U$, *and* $S \subseteq U^2$. *Assume that the complex algebra* $\mathfrak{A}$ *of* $\mathfrak{U}$ *is a relation algebra in which every atom is a functional element. Then we have:*

(i) $\mathfrak{U}$ *is a generalized Brandt groupoid.*

(ii) *If* $\mathfrak{A}$ *is simple, then* $\mathfrak{U}$ *is a Brandt groupoid.*

Proof. By 3. 8, 3. 9 and the hypothesis, $\mathfrak{A}$ is a complete atomistic relation algebra, with at least two different elements, in which every atom is a functional element. Hence, by 5. 6, $\mathfrak{A}$ is isomorphic to the complex algebra $\mathfrak{A}'$ of a system $\mathfrak{U}'$ which is a generalized Brandt groupoid; moreover, $\mathfrak{U}'$ is a Brandt groupoid in case $\mathfrak{A}$ is simple. From 3. 2 and 3. 8 we easily see that any two algebras in the wider sense are isomorphic whenever their complex algebras are isomorphic. Hence $\mathfrak{U}$ is isomorphic to $\mathfrak{U}'$; therefore $\mathfrak{U}$ is a generalized Brandt groupoid and, in case $\mathfrak{A}$ is simple, it is a Brandt groupoid.

As is seen from 4.19, the assumption that every atom in $\mathfrak{A}$ is functional can be replaced in the hypothesis of 5.7 by the assumption that R is a binary operation (on a subset of U^2 to U), and not only a ternary relation.

THEOREM 5.8. *For every relation algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

in which $0 \neq 1$ *the following two conditions are equivalent:*

(i) $\mathfrak{A}$ *is isomorphic to a proper relation algebra,*

(ii) $\mathfrak{A}$ *is isomorphic to a subalgebra of the complex algebra of a generalized Brandt groupoid.*

Also the following two conditions are equivalent:

(i′) $\mathfrak{A}$ *is isomorphic to a proper relation algebra on a set* U,

(ii′) $\mathfrak{A}$ *is isomorphic to a subalgebra of the complex algebra of a Brandt groupoid.*

Proof. Condition (i) implies (ii) by 4.31(i)(iii) and 5.6(i). Conversely, (ii) implies (i) by 4.31(i)(iii) and 5.5(i). In an analogous way we show that conditions (i′) and (ii′) are equivalent; we make use of the last part of 4.31 (concerning a modification of conditions 4.31(i)-(iii)), as well as of 5.5(ii) and 5.6(ii).

The following informal remarks concern an interesting class of generalized Brandt groupoids which can be referred to as having the *unicity property.* Generalized Brandt groupoids

$$\mathfrak{U} = \langle U, \cdot, I, {}^{-1} \rangle$$

of this class are characterized by the following condition:

(i) For any $x, z \in U$ there is at most one element $y \in U$ such that $x \cdot y \in U$ and $y \cdot z \in U$.

Another, equivalent formulation of this condition is

(ii) For any $x, y \in U$, if $x \cdot y \in U$ and $y \cdot x \in U$, then $y = x^{-1}$.

If V is a non-empty equivalence relation, I the set of all couples $\langle x, x \rangle \in V$, and if we define the operations $\cdot$ and ${}^{-1}$ by putting, for any couples $\langle x, y \rangle$, $\langle y, z \rangle \in V$,

$$\langle x, y \rangle \cdot \langle y, z \rangle = \langle x, z \rangle, \quad \langle x, y \rangle^{-1} = \langle y, x \rangle,$$

and by assuming that $\langle x, y \rangle \cdot \langle y', z \rangle$ does not exist in case $y \neq y'$, then, as is easily seen, the system

$$\mathfrak{U}(V) = \langle V, \cdot, I, {}^{-1} \rangle$$

is a generalized Brandt groupoid with the unicity property. Conversely, every generalized Brandt groupoid $\mathfrak{U}$ with the unicity property is isomorphic to a groupoid $\mathfrak{U}(V)$ constructed in the way just described. To obtain a function f mapping $\mathfrak{U}$ on a groupoid $\mathfrak{U}(V)$, we put for every element x in U

$$f(x) = \langle x \cdot x^{-1}, x^{-1} \cdot x \rangle.$$

By comparing these results with 5.6, we easily see that for every relation algebra $\mathfrak{A}$ the following two conditions are equivalent:

(i′) $\mathfrak{A}$ is isomorphic to the proper relation algebra constituted by all subrelations of a non-empty equivalence relation V, and

(ii′) $\mathfrak{A}$ is isomorphic to the complex algebra of a generalized Brandt groupoid $\mathfrak{U}$ with the unicity property.

As a consequence, we obtain by 5.8 that the complex algebra of an arbitrary generalized Brandt groupoid is isomorphic to a subalgebra of the complex algebra of a generalized Brandt groupoid with the unicity property. Either with the help of this result or in a more direct way we conclude that every generalized Brandt groupoid $\mathfrak{U}$ is a homomorphic image of a generalized Brandt groupoid $\mathfrak{U}'$ with the unicity property. In fact, we can take for $\mathfrak{U}'$ the system $\mathfrak{U}(V)$ defined above where V is the set of all couples $\langle x, y\rangle$ such that x, y, and $x \cdot y^{-1}$ are elements in U. The function g defined for every couple $\langle x, y\rangle \in V$ by the formula

$$g(x, y) = x \cdot y^{-1}$$

maps $\mathfrak{U}(V)$ homomorphically onto $\mathfrak{U}$.

All these remarks remain valid if we replace in them arbitrary generalized Brandt groupoids by Brandt groupoids, and arbitrary equivalence relations by relations of the form $V = W^2$ where W is an arbitrary set.

To conclude this section, we want to give some applications of our results to groups. A group may be considered as a system constituted by a non-empty set U, a binary operation $\cdot$ on U^2 to U, a distinguished element u of U (the unit or identity element), and a unary operation $^{-1}$ on U to U; the postulates which are to be satisfied by these notions are well known. We can, of course replace the element $u \in U$ by a set $I \subseteq U$ which is assumed to consist of just one element. Groups become then systems of the same type as Brandt groupoids, and we can state the following

THEOREM 5.9. *For*

$$\mathfrak{U} = \langle U, \cdot, I, ^{-1}\rangle$$

to be a group it is necessary and sufficient that $\mathfrak{U}$ *be a generalized Brandt groupoid in which* $\cdot$ *is an operation on* U^2 *to* U. *The condition remains necessary and sufficient if we omit in it the word "generalized."*

Proof. The theorem easily follows from the definition of a group and that of a (generalized) Brandt groupoid. The only thing which is perhaps not quite obvious is that in a generalized Brandt groupoid in which $\cdot$ is an operation on U^2 (and not only on a subset of U^2) to U the set I consists of just one element. To show this notice that I is non-empty by 5.1(v) and that, for any $x, y \varepsilon I$, we have $x \cdot x = x = x \cdot y$ by 5.2(i), and hence $x = y$ by 5.1(ii).

THEOREM 5.10. *The complex algebra of a group is a complete atomistic integral relation algebra in which every atom is a functional element.*[24]

Proof. By 4.16, 5.5(i), and 5.9, with the help of 3.2 and 3.8.

THEOREM 5.11. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ;, 1', \smile \rangle$$

be an atomistic integral relation algebra in which every atom is a functional element. Let U *be the set of all atoms of* $\mathfrak{A}$, *and let*

$$\mathfrak{U} = \langle U, ;, \{1'\}, \smile \rangle.$$

We then have:

(i) $\mathfrak{U}$ *is a group and* $\mathfrak{A}$ *is isomorphic to a subalgebra of the complex algebra of* $\mathfrak{U}$.

(ii) *If* $\mathfrak{A}$ *is complete, it is isomorphic to the complex algebra of* $\mathfrak{U}$.

Proof. By 4.17(i)(iv), 1' is an atom. Hence $\{1'\}$ coincides with the set of all atoms $u \leq 1'$, and therefore the system $\mathfrak{U}$ defined in the hypothesis coincides with the system $\mathfrak{U}$ of 5.6. Consequently, by 5.6(i), $\mathfrak{U}$ is a generalized Braandt groupoid. By 4.20, the operation $;$ in $\mathfrak{U}$ is an operation on U^2 to U. Hence, by 5.9, $\mathfrak{U}$ is a group. The remaining conclusions follow directly from 5.6(i) (ii).

THEOREM 5.12. *Let*

$$\mathfrak{U} = \langle U, R, I, S \rangle$$

be an algebra in the wider sense in which $R \subseteq U^3$, $I \subseteq U$, *and* $S \subseteq U^2$. *If the complex algebra* $\mathfrak{A}$ *of* $\mathfrak{U}$ *is an integral relation algebra in which every atom is a functional element, then* $\mathfrak{U}$ *is a group.*

[24] This is a result of J. C. C. McKinsey; see Jónsson-Tarski [2].

Proof. By 5.7, $\mathfrak{U}$ is a generalized Brandt groupoid. From 4.20, with the help of 3.2 and 3.8, we conclude that R is a binary operation on U^2 to U. The conclusion follows by 5.9.

Instead of assuming, in the hypothesis of 5.12, that $\mathfrak{A}$ is integral and that every atom in $\mathfrak{A}$ is a functional element, it suffices to assume that $\mathfrak{U}$ is an algebra, and not only an algebra in the wider sense. Even the (weaker) assumption that R is a binary operation on U^2 to U proves to be sufficient. Compare the analogous remark following 5.7.

The results stated in 5.10 and 5.11 can partly be extended to arbitrary Boolean algebras with operators and to complex algebras of arbitrary algebras (not in the wider sense). A Boolean algebra with operators

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle$$

in which $0 \neq 1$ is called *integral* if, for each of the functions f_ξ with the rank m_ξ and for every sequence $x \in A^{m_\xi}$ the formula $f_\xi(x) = 0$ implies that $x_j = 0$ for some $j < m_\xi$. (In case $\mathfrak{A}$ is a relation algebra with $0 \neq 1$, this condition is automatically satisfied both by $1'$ treated as an operation with the rank 0 and by $^\smile$, and has to be postulated only for $;$.) The condition that $\mathfrak{A}$ is integral is clearly necessary for $\mathfrak{A}$ to be isomorphic to the complex algebra of some algebra. This condition is, however, not sufficient. On the contrary, the following condition—when combined with completeness, atomisticity, and the condition that $0 \neq 1$—is necessary and sufficient for $\mathfrak{A}$ to be isomorphic to the complex algebra of some algebra: for each of the functions f_ξ with the rank m_ξ and for every sequence $x \in U^{m_\xi}$ where U is the set of all atoms in $\mathfrak{A}$ we have $f_\xi(x) \in U$. It is seen from 4.3(xii), 4.17(i)(iv), and 4.20 that the latter condition, when applied to an atomistic relation algebra $\mathfrak{A}$, is equivalent to the one occurring in 5.10 and 5.11, i. e., to the condition that $\mathfrak{A}$ is integral and that every atom in $\mathfrak{A}$ is a functional element.

By 4.18(i) and 5.10, every relation algebra which is a subalgebra of the complex algebra of a group is integral. The question whether, conversely, every integral relation algebra is isomorphic to a subalgebra of the complex algebra of a group is still open. The answer is not known even for those integral algebras which are isomorphic to proper relation algebras.

A REMARK ON BOOLEAN ALGEBRAS WITH OPERATORS.*

By Hugo Ribeiro.

Bjarni Jónsson and Alfred Tarski, in their paper *Boolean algebras with operators* (**73**, 891 and **74**, 127 of this Journal), call a Boolean algebra B a regular subalgebra of a Boolean algebra A if A is complete and atomistic and B is a subalgebra of A for which: i) if I is an arbitrary set and if the elements $x_i \varepsilon B$ with $i \varepsilon I$ are such that $\sum_{i \varepsilon I} x_i = 1$, then there exists a finite subset J of I such that $\sum_{i \varepsilon J} x_i = 1$, ii) if u and v are distinct atoms of A then there exists an element $b \varepsilon B$ such that $u \leqq b$ and $v \cdot b = 0$ (Definition 1.19). The set C of all "closed" elements of A is then defined as the set of all elements $x \varepsilon A$ such that $x = \prod_{x \leq y \varepsilon B} y$ (Definition 1.20); and A^m, B^m, C^m designate the sets of all m-termed sequences, $x = \langle x_0, \cdots, x_{m-1} \rangle$, of elements of A, B, C respectively. Furthermore, a function f on B^m to B is called monotonic if given two sequences $x \leq y \varepsilon B^m$ (that is $x_i \leq y_i \varepsilon B$ for any $i = 0, \cdots, m-1$) we always have $f(x) \leq f(y)$, additive if given any $j < m$ and $x, y \varepsilon B^m$ such that $x_p = y_p$ whenever $j \neq p < m$ we always have $f(x+y) = f(x) + f(y)$ (Definition 1.1); f^+/C^m designates the restriction of the function f^+ to C^m and by the composition $f[g_0, \cdots, g_{m-1}]$ of f, on B^m to B, with $g_0, \cdots, g_{m-1}$, on B^n to B, it is understood the function h on B^n to B such that $h(x) = f(g_0(x), \cdots, g_{m-1}(x))$ whenever $x \varepsilon B^n$. Finally, to any function f on B^m to B an extension, f^+, on A^m to A is defined by $f^+(x) = \sum_{x \geq y \varepsilon C^m} \prod_{y \leq z \varepsilon B^m} f(z)$ for any $x \varepsilon A^m$ (Definition 2.1), and it is shown (as an immediate consequence of Theorem 2.10) that if an equation involving additive functions on B^m to B is identically satisfied, then the corresponding equation involving their extensions is also identically satisfied. Such a statement is also true (Theorem 2.11) of certain implications between two equations, and it yields several interesting results.

The purpose of the present note is to give a direct proof of an extension of that Jónsson-Tarski's Theorem 2.10. This extension (Theorem II) consists in getting the conclusion under a weaker hypothesis on the functions $g_0, \cdots, g_{m-1}$, namely the monotonicity instead of membership in the set ϕ.

* Received March 31, 1950.

Otherwise our statement (Theorem II) is as Jónsson-Tarski's Theorem 2.10. (It must be pointed out that in that same paper it is also shown: Theorem 2.10 does not hold with the hypothesis that f is monotonic even when $g_0, \cdots, g_{m-1}$ are additive, and on the other hand, it holds whenever f is monotonic and $g_0, \cdots, g_{m-1}$ are "identity functions" (Theorem 2.9)). Throughout our proof we shall make free use of many of the terminology and notation in Jónsson-Tarski's paper, and we shall continue to refer to its definitions and theorems by using the reference numbers therein.[1]

THEOREM I. *Let $B = \langle B_0, +, 0, \cdot, 1\rangle$ be a regular subalgebra of a Boolean algebra $A = \langle A_0, +, 0, \cdot, 1\rangle$ and let m and n be positive integers. Then, if f is an additive function on B^m to B and $g_0, \cdots, g_{m-1}$ are monotonic functions on B^n to B we have*

$$(f[g_0, \cdots, g_{m-1}])^+ = f^+[g_0^+, \cdots, g_{m-1}^+].$$

Proof. First we remark that f^+/C^m is on C^m to C and g_j^+/C^n $(j = 0, \cdots, m-1)$ are on C^n to C. The inclusion

1) $f^+[g_0^+, \cdots, g_{m-1}^+] \leq (f[g_0, \cdots, g_{m-1}])^+$ is easily checked: Using the definition of composition, 2.1 and 2.2, the hypothesis on f together with 2.4, and then the remark that $x \geq y^0, \cdots, y^{m-1} \in C^n$ implies $x \geq y^0 + \cdots + y^{m-1} \in C^n$ together with the monotonicity of f^+ and g_j^+ $(j = 0, \cdots, m-1)$, we have for every $x \in A^n$

$$\begin{aligned} f^+[g_0^+, \cdots, g_{m-1}^+](x) &= f^+(g_0^+(x), \cdots, g_{m-1}(x)) \\ &= f^+\Big(\sum_{x \geq y^0 \in C^n} g_0^+(y^0), \cdots, \sum_{x \geq y^{m-1} \in C^n} g_{m-1}^+(y^{m-1})\Big) \\ &= \sum_{x \geq y^0 \in C^n} \cdots \sum_{x \geq y^{m-1} \in C^n} f^+(g_0^+(y^0), \cdots, g_{m-1}^+(y^{m-1})) \\ &\leq \sum_{x \geq y \in C^n} f^+(g_0^+(y), \cdots, g_{m-1}^+(y)). \end{aligned}$$

By 2.2 the last sum is $\sum_{x \geq y \in C^n} \prod_{\langle g_0^+(y), \ldots, g_{m-1}^+(y)\rangle \leq z \in B^m} f(z)$ and it is included in $\sum_{x \geq y \in C^n} \prod_{y \leq z \in B^n} f(g_0(z), \cdots, g_{m-1}(z))$, since every factor of each product of this sum is a factor of the corresponding product of the above sum: f is monotonic, and $\langle g_0^+(y), \cdots, g_{m-1}^+(y)\rangle \leq \langle g_0(z), \cdots, g_{m-1}(z)\rangle \in B^m$ whenever $y \leq z \in B^n$ because of the monotonicity of g_j^+ $(j = 0, \cdots, m-1)$. Now

[1] The results of the present note were originated from and reported to Professor Tarski's seminar on Topics in algebra and metamathematics at the University of California, Berkeley.

$$\sum_{x \geq y \varepsilon C^n} \prod_{y \leq z \varepsilon B^n} f(g_0(z), \cdots, g_{m-1}(z))$$
$$= \sum_{x \geq y \varepsilon C^n} \prod_{y \leq z \varepsilon B^n} f[g_0, \cdots, g_{m-1}](z) = (f[g_0, \cdots, g_{m-1}])^+(x),$$

by the definition of composition and then by 2. 1.

From this proof of 1) it follows that, for every $x \varepsilon C^n$ such an inclusion holds even if f is monotonic not additive. In this case the sequence of equalities and inclusions yielding 1) will, essentially, begin after the first inclusion above, and there the additivity of f does not play any role.

Next, we prove for f monotonic (not necessarily additive) the inclusion

2) $$(f([g_0, \cdots, g_{m-1}])^+ \leq f^+[g_0^+, \cdots, g_{m-}^+].$$

From the definition of composition, 2. 1 and 2. 2 it is clear that it will be sufficient to show that for every $y \varepsilon C^n$ we have

$$(f[g_0, \cdots, g_{m-1}])^+(y) \leq f^+(g_0^+(y), \cdots, g_{m-1}^+(y)).$$

By 2. 2 and, again by the definition of composition, this inclusion will be established, for any $y \varepsilon C^n$, if we prove that to each factor of

$$\prod_{\langle g_0^+(y), \ldots, g_{m-1}^+(y)\rangle \leq z' \varepsilon B^m} f(z')$$

there is at least one factor of $\prod_{y \leq z'' \varepsilon B^n} f(g_0(z''), \cdots, g_{m-1}(z''))$ which is included in it.

Since f is monotonic it is now sufficient to show that to each $z' \varepsilon B^m$ for which $\langle g_0^+(y), \cdots, g_{m-1}^+(y)\rangle \leq z'$, there is some $z'' \varepsilon B^n$ having the properties I) $y \leq z''$, II) $\langle g_0(z''), \cdots, g_{m-1}(z'')\rangle \leq z'$.

To do this let $z' = \langle z'_0, \cdots, z'_{m-1}\rangle \varepsilon B^m$ and let us remark that, by 2. 2, our hypothesis means

$$\prod_{y \leq z \varepsilon B^n} g_j(z) \leq z'_j \varepsilon B \qquad (j = 0, \cdots, m-1).$$

First, we have that *for any* $j = 0, \cdots, m-1$ *there is* $z''^j \varepsilon B^n$ *such that at same time* $y \leqq z''^j$ *and* $g_j(z''^j) \leqq z'_j$. This is true since z'_j being open and including a product of closed elements, it will include (by 1. 21, (iv)) some finite product $\prod_{k=0}^{r_j} g_j(z^k)$ of such closed (and open) factors:

$$\prod_{k=0}^{r_j} g_j(z^k) \leq z'_j \qquad (k = 0, \cdots, r_j)$$

with $y \leq z^k \varepsilon B^n$. Now, putting $z''^j = \prod_{k=0}^{r_j} z^k$, we will have not only $y \leq z''^j \varepsilon B$

but also $g_j(z''^j) \leq z'_j$, since $g_j(z''^j) \leq \prod_{k=0}^{r_j} g_j(z^k)$ because of the monotonicity of g_j. As second and final step it is easily seen that $z'' = \prod_{j=0}^{m-1} z''^j$ is an element of B^n having the properties I) and II) above: $y \leq z''$ since $y \leq z''^j$ for every $j = 0, \cdots, m-1$; and $g_j(z'') \leq z'_j$ for every $j = 0, \cdots, m-1$, since g_j being monotonic we have $g_j(z'') \leq g_j(z''^j) \leq z'_j$.

The proof of 2) is now complete. From 1) and 2) the theorem follows.

Remark. As a consequence of the preceding proof of 2) and of the comment at the end of the proof of 1) we have

$$(f[g_0, \cdots, g_{m-1}])^+(y) = f^+[g_0^+, \cdots, g_{m-1}^+](y)$$

whenever $f, g_0, \cdots, g_{m-1}$ are monotonic and $y \in C^n$.

THEOREM II. *Let $B = \langle B_0, +, 0, \cdot, 1\rangle$ be a regular subalgebra of a Boolean algebra $A = \langle A_0, +, 0, \cdot, 1\rangle$, let m and n be positive integers and let ϕ be the smallest set having the two properties:*

i) *to include all additive functions on B^t to B for any t (integer positive)*

ii) *to be closed in respect to the operation of composition (of functions).*

Then, if $g_0, \cdots, g_{m-1}$ are monotonic functions on B^n to B and $f \in \phi$ is a function on B^m to B, we have

$$(f[g_0, \cdots, g_{m-1}])^+ = f^+[g_0^+, \cdots, g_{m-1}^+].$$

Proof. Remark first that the operation of composition is associative and that for a function f on B^m to B to verify the hypothesis it is necessary (and sufficient) that non negative integers k and $r_0 = 0, r_1, \cdots, r_k$ exist such that

$$f = h_0^0[h_0^1, \cdots, h_{r_1}^1] \cdots [h_0^k, \cdots, h_{r_k}^k]$$

for some additive functions h_j^k $(j = 0, \cdots, r_k)$ on B^m to B and h_j^i $(i = 0, \cdots, k-1; j = 0, \cdots, r_i)$ on $B^{r_{i+1}}$ to B.

Put $h = h_0^0$. If we have $k = 0$ in the equality above, then f is just the additive function h, and we have the desired conclusion from Theorem I. We prove by induction for $k \neq 0$. Putting

$$f'_j = h_j^1[h_0^2, \cdots, h_{r_2}^2] \cdots [h_0^k, \cdots, h_{r_k}^k] \qquad (j = 0, \cdots, r_1)$$

we have

$$(f'_j[g_0, \cdots, g_{m-1}])^+ = f_j'^+[g_0^+ \cdot \cdots, g_{m-1}] \qquad (j = 0, \cdots, r_1)$$

as induction hypothesis. On the other hand, since f'_j and also $f'_j[g_0, \cdots, g_{m-1}]$ $(j = 0, \cdots, r_1)$ are monotonic and h is additive, we have

$$h^+[f'_0{}^+, \cdots, f'_{r_1}{}^+] = (h[f'_0, \cdots, f'_{r_1}])^+$$

and also

$$\begin{aligned}(h[f'_0[g_0, \cdots, g_{m-1}], \cdots, f'_{r_1}[g_0, \cdots, g_{m-1}]])^+ \\ = h^+[(f'_0[g_0, \cdots, g_{m-1}])^+, \cdots, (f'_{r_1}[g_0, \cdots, g_{m-1}])^+],\end{aligned}$$

by Theorem I. Hence,

$$\begin{aligned}(f[g_0, \cdots, g_{m-1}])^+ &= ((h[f'_0, \cdots, f'_{r_1}])[g_0, \cdots, g_{m-1}])^+ \\ &= (h[f'_0[g_0, \cdots, g_{m-1}], \cdots, f'_{r_1}[g_0, \cdots, g_{m-1}]]^+ \\ &= h^+[(f'_0[g_0, \cdots, g_{m-1}])^+, \cdots, (f'_{r_1}[g_0, \cdots, g_{m-1}])^+] \\ &= h^+[f'_0{}^+[g_0{}^+, \cdots, g_{m-1}{}^+], \cdots, f'_{r_1}{}^+[g_0{}^+, \cdots, g_{m-1}{}^+]] \\ &= (h^+[f'_0{}^+, \cdots, f'_{r_1}{}^+])[g_0{}^+, \cdots, g_{m-1}{}^+] \\ &= (h[f'_0, \cdots, f'_{r_1}])^+[g_0{}^+, \cdots, g_{m-1}{}^+] = f^+[g_0{}^+, \cdots, g_{m-1}{}^+].\end{aligned}$$

Thus Theorem II is proved.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

CRITÈRES DE COMPACITÉ DANS LES ESPACES FONCTIONNELS GENERAUX.*

By A. Grothendieck.

1. Introduction.[1] Outre la notion usuelle de compacité (cf. N. Bourbaki [1]), qui est apparue come étant la seule qui soit vraiment fondamentale, on rencontre néanmoins au moins deux autres notions étroitement apparentées, mais de "caractère dénombrable," et qui se sont révélées indispensables dans plusieurs questions qui à priori n'impliquent aucune considération de dénombrabilité.

Nous dirons qu'une partie A d'un espace topologique est *relativement semi-compacte* (resp. *semi-compacte*) si toute suite extraite de A admet une valeur d'adhérence (resp. qui appartienne à A). A sera dite *strictement relativement semi-compacte* (resp. *strictement semi-compacte*) si de toute suite extraite de A on peut extraire une suite converge (resp. qui converge vers un élément de A).

Dans le présent travail, nous étudions des cas étendus où la semi-compacité relative entraine déjà la compacité relative ou la stricte semi-compacité relative. Nous nous y plaçons surtout dans des espaces du type $C_{\mathfrak{S}}(E, F)$, espace des applications continues d'un espace topologique E dans un espace uniforme séparé F, muni de la topologie de la convergence uniforme sur un ensemble $\mathfrak{S}$ de parties de E recouvrant E (pour ces notions fondamentales d'Analyse Fonctionnelle, cf. N. Bourbaki [3]). Les résultats obtenus valent manifestement pour les sous-espaces fermés de tels espaces, ce qui permet de les appliquer à des espaces vectoriels localement convexes généraux.

Dans 2, nous donnons quelques généralités sur les diverses notions de compacité envisagées, destinées surtout à prémunir le lecteur contre certaines

---

* Received December 27, 1950.

[1] Outre le théorème 5 et la proposition 7 du présent travail, citons notamment encore deux propositions s'appuyant de façon essentielle sur le théorème d'Eberlein (cf. proposition 2 ci-dessous): 1) l'enveloppe convexe fermée d'une partie faiblement relativement compacte d'un espace de Banach (par exemple) est faiblement compacte; 2) Le produit de deux fonctions faiblement presque-périodiques sur un semi-groupe est faiblement presque-périodique (cf. [6]); plus généralement, si E est une algèbre normée complète s'identifiant à l'espace des fonctions complexes continues sur un espace compact, le produit de deux parties faiblement compactes A et B de E (ensemble des xy avec $x \varepsilon A$ et $y \varepsilon B$) est faiblement relativement compact.

erreurs assez naturelles, plutôt que de repéter les développements bien connus et triviaux sur le sujet (tels que: l'image continue d'un espace semi-compact est semi-compact, etc.). Dans 3. nous étudions des cas où la semi-compacité relative entraine la compacité relative, le resultat le plus important est le théorème 2 (et le théorème 1 qui est un corollaire) ; nous y donnons en même temps un critère de compacité relative qui semble à priori encore bien plus faible que la semi-compacité relative. En outre, M. J. Dieudonné a bien voulu me communiquer un autre cas non trivial et très simple où la semi-compacité relative entraine la compacité relative (théorème 3), critère qui ne sera pas essentiel par la suite mais a son intérêt propre dans l'ordre d'idées de ce travail. Dans 4, nous appliquons les résultats obtenus aux espaces localement convexes, en donnant notamment à un classique théorème d'Eberlein pour les espaces de Banach (généralisé par J. Dieudonné et L. Schwartz [5] aux espaces ($\mathfrak{F}$)) toute la généralité qui lui appartient. (Ce théorème a été d'ailleurs le point de départ du présent travail). Dans 5. nous établissons un cas non classique ou la semi-compacité relative entraine la *stricte* semi-compacité relative (th. 4) ; ce résultat est d'ailleurs essentiel pour la suite (th. 5) ; la proposition 5 se réduit à une systématisation de réflexions classiques. Nous appliquons ensuite les resultats précédents à la détermination des parties faiblement relativement compactes de l'espace de Banach $C^{\infty}(E)$ de toutes les fonctions continues et bornées sur un espace topologique E; le critère obtenu donne par exemple immédiatement le résultat suivant. Une fonction faiblement presque périodique à gauche sur un semi-groupe est aussi faiblement presque périodique à droite. Enfin dans 7. nous généralisons le théorème 6 pour obtenir dans les espaces localement convexes un critère de relative compacité faible, approfondissant de beaucoup les résultats de 4, et qui ne semble pas connu même pour les espaces de Banach.

2. Généralités. Il n'est peut-être pas inutile de rappeler quelles implications on peut ou ne peut pas affirmer entre les diverses notions de compacité envisagées, et quelles simplifications se produisent dans quelques cas classiques. Il est évident que la compacité et la semi-compacité stricte entraînent chacune la semi-compacité, de même pour les notions " relatives " correspondantes. Mais on n'a dans le cas général aucune autre implication entre ces trois couples de notions, car il est bien connu qu'un espace strictement semi-compact peut être non compact (exemple: espace des nombres ordinaux de seconde classe, avec la topologie usuelle) et un espace compact peut ne pas être strictement semi-compact (exemple: produit topologique d'une famille non dénombrable d'intervalles compacts).—D'autre part, il est évident que chacune des trois notions de compacité entraine la notion " relative " correspondante, et que la réciproque est tout à fait fausse. On fera attention ici que pour qu'une partie A d'un espace E soit relativement compacte, il faut et il suffit par définition que son adhérence soit compacte, mais qu'il n'est est plus de même pour la semi-compacité relative et la stricte semi-compacité relative. La condition est évidemment encore suffisante, mais on peut trouver une partie A strictement semi-compacte d'un espace séparé E, dont l'adhérence ne soit pas même semi-compacte. En d'autres termes, il existe un espace séparé non semi-compact E dans lequel une partie strictement semi-compacte A soit dense. Soit en effet X un espace séparé qui soit strictement semi-compact et localement compact mais non compact (par exemple l'espace des nombres ordinaux de seconde classe), a son " point à l'infini," $Y = X \cup (a)$. Pour tout entier naturel n, soit Y_n un exemplaire homéomorphe de Y (X_n correspondant à X et a_n à a) ; supposons les Y_n disjoints et soit b un élément qui n'appartienne à aucun des Y_n. Sur l'ensemble $E = (b) \cup \bigcup_n Y_n$, considérons la topologie dont les ouverts sont les parties qui coupent chaque Y_n suivant un ouvert, et qui, s'ils contiennent b, contiennent aussi les X_n à partir d'un rang assez élevé. On vérifie trivialement les axiomes des ouverts (Bourbaki (1)), et que E est séparé; E n'est pas semi-compact, car il est manifeste que la suite $(a_n)_n$ n'a pas de point adhérent. D'autre part, $A = (b) \cup \bigcup_n X_n$ est partout dense et strictement semi-compact, comme on vérifie aussitôt.

Rappelons enfin que dans un espace métrique, les trois notions de compacité sont équivalentes, ainsi que les notions " relatives " correspondantes. Un autre résultat intéressant, qui nous sera utile par la suite, est le théorème d'A. Weil [10] ; une partie relativement semi-compacte d'un espace uniforme séparé est précompacte. En particulier, dans un espace uniforme séparé et complet, compacité (relative) et semi-compacité (relative) sont la même chose.

La topologie faible d'un espace de Banach ou d'un espace localement convexe quelconque est un exemple d'un topologie en général ni complète ni métrisable, et où les critères de relative compacité qu'on vient de rappeler ne s'appliquent pas tels quels. Plus généralement, il en est ainsi dans les espaces d'applications continues d'un espace topologique dans un autre, muni de la topologie de la convergence simple par exemple. Pourtant un théorème d'Eberlein pour les espaces de Banach, et les résultats de G. Köthe sur ses " espaces parfaits " (cf. G. Köthe [8]) montrent que dans ces espaces, munis de la topologie faible, on a encore identité entre parties relativement semi-compactes et relativement compactes. D'autre part, dans les espaces de Banach encore, un théorème de Šmulian (généralisé aux espaces ($\mathfrak{F}$) dans [5]) affirme l'identité pour la topologie faible entre parties strictement relativement semi-compactes et parties relativement compactes. Ce sont ces résultats qui nous ont guidé et que nous allons généraliser et préciser.

3. Semi-compacité et compacité. Soit E un espace topologique, F un espace uniforme séparé; nous désignons par $C(E,F)$ l'espace des applications continues de E dans F, par $\mathfrak{F}(E,F)$ l'espace de toutes les applications de E dans F, et, si $\mathfrak{S}$ est un ensemble de parties de E, par $\mathfrak{T}_{\mathfrak{S}}$ la structure uniforme sur $C(E,F)$ et $\mathfrak{F}(E,F)$ de la convergence uniforme sur les éléments de $\mathfrak{S}$ (cf. [3]); munis de cette structure les espaces précédents seronts désignés par $C_{\mathfrak{S}}(E,F)$ resp. $\mathfrak{F}_{\mathfrak{S}}(E,F)$. Si F est séparé et si $\mathfrak{S}$ recouvre E (ce que nous supposons par la suite) ces espaces sont séparés, et si de plus F est complet, il en est même de $\mathfrak{F}_{\mathfrak{S}}(E,F)$, mais en général $C_{\mathfrak{S}}(E,F)$ n'est pas complet. On vérifie aussitôt que si un filtre de Cauchy dans $\mathfrak{F}_{\mathfrak{S}}(E,F)$ converge pour la topologie de la convergence simple, il converge pour $\mathfrak{T}_{\mathfrak{S}}$, d'où suit que pour qu'une partie A de $C_{\mathfrak{S}}(E,F)$ ait une adhérence complète dans cet espace, il faut et il suffit que tout filtre de Cauchy sur A converge en chaque point vers une application *continue* de E dans F. Alors l'adhérence $\bar{A}$ de A pour $\mathfrak{T}_{\mathfrak{S}}$ sera à fortiori complète pour toute $\mathfrak{T}_{\mathfrak{S}'}$, avec $\mathfrak{S}' \supset \mathfrak{S}$. En particulier si A a une adhérence complète dans l'espace $C_s(E,F)$ muni de la structure $\mathfrak{T}_s$ de la convergence simple, il en sera de même à fortiori pour toute $\mathfrak{T}_{\mathfrak{S}}$. A fortiori, si A est relativement compacte dans $C_s(E,F)$, l'adhérence de A dans $C_{\mathfrak{S}}(E,F)$ est complète quel que soit l'ensemble de parties $\mathfrak{S}$.

Enfin, remarquons encore que le théorème de Tychonoff donne immédiatement: Pour qu'une partie de $C_s(E,F)$ soit relativement compacte, il faut et il suffit que 1°) elle le soit dans le produit topologique $\mathfrak{F}_s(E,F)$, c'est-à-dire que pour tout $x \in E$, l'ensemble des $f(x)$, où $f \in A$, soit relativement compact dans F; et 2°) que l'adhérence de A dans $C_s(E,F)$ soit la même que dans $\mathfrak{F}_s(E,F)$, c'est-à-dire que toute application de E dans F qui est limite simple d'applications éléments de A, soit continue.

Ces remarques interviennent dans diverses questions d'Analyse Fonctionnelle, et seront essentielles pour la compréhension de la suite.

THÉORÈME 1. *Soit E un espace semi-compact, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Si dans F toute partie relativement semi-compacte est relativement compacte (en particulier, si F est complet), alors il en est de même dans l'espace $C_{\mathfrak{S}}(E,F)$.*

Toute partie relativement semi-compacte A de $C_{\mathfrak{S}}(E,F)$ est précompacte (cf. ci-dessus 2.); if suffit de montrer que son adhérence est complète, et a fortiori, d'après nos remarques précédentes, que A est relativement compact pour la topologie de la convergence simple. Comme A est évidemment relativement semi-compact pour cette dernière topologie (puisque $\mathfrak{S}$ recouvre E) on est ramené au cas de la topologie $\mathfrak{T}_s$. Mais ce cas est inclus dans le théorème-clef suivant:

THÉORÈME 2. *Soit E un espace semi-compact, F un espace complètement régulier, A une partie de $C_s(E,F)$, E_1 une partie dense de E.*

1°) *Si dans F toute partie relativement semi-compacte est relativement compacte, alors les conditions suivantes sur A sont toutes équivalentes:*

a) *A est relativement compact;*

b) *A est relativement semi-compact;*

c) *pour toute suite (f_n) extraite de A et toute suite (x_i) extraite de E_1, il existe une application continue f de l'adhérence K de l'ensemble des x_i dans F, telle que pour tout $x \in K$, $f(x)$ soit adhérent à la suite des $f_n(x)$. Et pour tout $x \in \complement E_1$, l'ensemble des $f(x)$ avec $f \in A$ est relativement semi-compact;*

d) *pour toute suite (f_n) extraite de A et toute suite (x_i) extraite de E_1, il existe un $X \in F$ qui soit point doublement adhérent à la suite double $(f_n(x_i))$ (par quoi nous entendons que tout voisinage de X rencontre une infinité de lignes et une infinité de colonnes de la suite double chacune en une infinité de termes). Et pour tout $x \in \complement E_1$, l'ensemble des $f(x)$ avec $f \in A$ est relativement semi-compact.*

2°) *De toutes façons (sans plus faire sur F la restriction de la première partie de l'énoncé) chacune des conditions qui précèdent est suffisante pour assurer que toute application de E dans F qui est limite simple d'applications éléments de A est continue; les deuxièmes parties des conditions* c) *et* d) *peuvent être omises.*

Enfin, moyennant la première partie de la condition d), *même si on ne suppose plus que E est semi-compact, toute application de E dans F qui est limite simple d'applications éléments de A est continue.*

Démonstration. On a de toutes façons manifestement a)$\rightarrow$b)$\rightarrow$c); montrons que si E est semi-compact, c) entraine d); il suffit de montrer que la première partie de la condition c) entraine la première partie de la condition d). Soit en effet, avec les notations de d), $x_0 \in E$ adhérent à la suite (x_i) et soit f l'application stipulée dans c), relative aux suites (x_i) et (f_n); je dis que $f(x_0)$ est doublement adhérent à la suite double $(f_n(x_i))$. En effet, s'il existait un voisinage ouvert V de $f(x_0)$ tel que sauf pour un nombre fini d'indices i l'on ait "$f_n(x_i) \varepsilon \complement V$ pour $n \geqq n_0(i)$," on aurait„ sauf pour un nombre fini d'indices: $f(x_i) \varepsilon \complement V$, d'où $f(x_0) \varepsilon \complement V$ ce qui est absurde; et s'il existait un voisinage ouvert V de $f(x_0)$ tel que sauf pour un nombre fini d'indices n l'on ait "$f_n(x_i) \varepsilon \complement V$ pour $i \geqq i_0(n)$," on aurait sauf pour un nombre fini d'indices $f_n(x_0) \varepsilon \complement V$, d'où $f(x_0) \varepsilon \complement V$, ce qui est encore absurde.— Pour prouver la première partie du théorème, tout revient donc à prouver que d) entraine a). Mais de d) resulte manifestement que pour *tout* $x \varepsilon E$ l'ensemble des $f(x)$, avec $f \varepsilon A$ est une partie relativement semi-compacte de F, donc relativement compacte en vertu de l'hypothèse sur F. En tenant compte d'une remarque faite plus haut, tout revient donc à montrer que toute application de E dans F qui est limite simple d'applications éléments de A est continue. Cela est inclus dans la deuxième partie du théorème, cette deuxième partie revenant manifestament à prouver que si E est un espace topologique quelconque, et F complètement régulier, alors la première condition énoncée dans d) est suffisante pour assurer que toute application f de E dans F qui est limite simple d'applications éléments de A est continue (E_1 désignant une partie dense dans E).

F étant complèment régulier, sa topologie peut être considérée comme la moins fine de celles qui rendent continues certaines fonctions numériques ϕ_i sur F ([2], page 11, proposition 4). On voit alors qu'on peut se ramener au cas où F est la droite numérique, la continuité de f équivalant en effet à la continuité de chacun des fonctions numériques $\phi_i \circ f$ sur E (d'autre part $\phi_i \circ f$ est limite simple de fonctions $\phi_i \circ g$ où g parcourt A, et l'ensemble de ces $\phi_i \circ g$ jouit manifestement des propriétés envisagées pour A lui-même). Supposons donc que F soit la droite numérique; on sait que pour démontrer la continuité de f, il suffit de montrer que l'on a $\lim\limits_{\substack{x \to x_0 \\ x \varepsilon E_1}} f(x) = f(x_0)$ pour tout $x_0 \varepsilon E$ ([1] page 38, th. 1). Nous démontrerons cette relation par l'absurde, en reprenant une idée d'Eberlein. Supposons donc qu'il existe un $x_0 \varepsilon E$ et un $\alpha > 0$ tels que pour tout voisinage V de x_0, il existe un $x \varepsilon V \cap E_1$ tel que $|f(x) - f(x_0)| \geqq \alpha$. On pourrait alors par récurrence construire deux suites d'éléments de A et de E_1 respectivement, (f_i) et (x_i), telles que l'on ait (on suppose les suites construites déjà jusqu'aux termes de rang $n-1$):

a) $|f_n(x_i) - f(x_i)| \leqq 1/n$ $(0 \leqq i \leqq n-1)$ (cela est possible, f étant limite simple d'éléments de A).

b) $|f_i(x_n) - f_i(x_0)| \leqq 1/n$ $(0 \leqq i \leqq n)$.

c) $|f(x_n) - f(x_0)| \geqq \alpha$ (ce que est encore possible par hypothèse).

Il existe un point z doublement adhérent à la suite double $(f_i(x_j))$, et comme pour i constant la suite $(f_n(x_i))$ tend vers $f(x_i)$ en vertu de a) et la suite $(f_i(x_n))$ vers $f_i(x_0)$ en vertu de b), z est adhérent à chancune des suites $(f(x_n))$ et $(f_n(x_0))$. Or, en vertu de la première des inégalités a), la deuxième suite tend vers $f(x_0)$; on a donc $z = f(x_0)$, et $f(x_0)$ serait valeur d'adhérence de la première suite, contrairement aux inégalités c). CQFD.

COROLLAIRE 1. *L'énoncé du théorème* 1 *reste valable si on suppose* E *localement compact, ou métrique, et plus généralement si toute application de* E *dans* F *dont les restrictions aux parties semi-compactes de* E *sont continues, est continue.*

En effet, on se ramène évidemment à montrer que si A est relativement semi-compact pour la topologie de la convergence simple, toute limite simple d'applications éléments de A est continue, ce qui est déjà une conséquence du théorème 1.

Remarque 1. La démonstration du théorème 2 met en évidence que si on suppose la topologie de F définie comme la moins fine des topologies rendant continues certaines applications ϕ_i de F dans des espaces complètement réguliers F_i, alors les critères énoncés dans le théorème 2 équivalent à ceux qu'on en déduit en supposant que les hypothèses envisagées sont vérifiées, non pour A lui-même à priori, mais pour chacun des ensembles $A_i \subset C(E, F_i)$ (où pour tout i, on désigne par A_i l'ensemble des $\phi_i \circ f$ où f parcourt A).

Remarque 2. Supposons que la suite double (x_{ij}) prenne ses valeurs dans un espace métrique F, et y soit relativement compacte. Alors on vérifie que la non-existence d'un point doublement adhérent à la suite double implique l'existence d'une "suite double extraite" $(x_{i_\alpha j_\beta}) = (y_{\alpha,\beta})$, telle que $\lim_\alpha . \lim_\beta . y_{\alpha\beta}$ et $\lim_\beta . \lim_\alpha . y_{\alpha\beta}$ existent tous deux et soient distincts. En effet, l'application du procédé diagonal permet de construire une suite d'indices (i_α) telle que $\lim_{\alpha\to\infty} . x_{i_\alpha,j}$ existe pour tout j. Une seconde application du procédé diagonal permet d'obtenir une suite d'indices j_β telle que $\lim_{\beta\to\infty} . x_{i_\alpha,j_\beta}$ existe pour tout i_α, et que $\lim_\beta . (\lim_\alpha . x_{i_\alpha,j_\beta})$ existe. Enfin on peut supposer en extrayant encore au besoin une suite partielle de la suite (i_α), que $\lim_\alpha . (\lim_\beta . x_{i_\alpha,j_\beta})$ existe. Mais les deux limites doubles $\lim_\alpha . (\lim_\beta . x_{i_\alpha,j_\beta})$ et $\lim_\beta . (\lim_\alpha . x_{i_\alpha,j_\beta})$ ne peuvent être égales, car leur valeur commune serait manifestement un point doublement adhérent à la suite double (x_{i_α,j_β}).

D'autre part, une suite double telle que $\lim_i . (\lim_j . x_{i,j})$ et $\lim_j . (\lim_i . x_{ij})$

existent et soient distincts n'a manifestement pas de point doublement adhérent (car un tel point devrait être identique à chacune de ces limites doubles). Il suit aussitôt le

COROLLAIRE 2. *Soit E un espace semi-compact, F un espace métrique, A un ensemble d'applications continues de E dans F tel que l'ensemble des $f(x)$ où $f \varepsilon A$ et $x \varepsilon E$, soit relativement compact. Pour que A soit relativement compact dans $C_s(E, F)$, il faut et il suffit qu'il n'existe pas de suite (x_i) extraite de E et de suite (f_j) extraite de A, telles que $\lim_i . \lim_j . f_j(x_i)$ et $\lim_j . \lim_i . f_j(x_i)$ existent tous deux et soient distincts. La condition subsiste si on assujettit la suite (x_{ij}) à être extraite d'une partie partout dense fixe E_1 de E. Et cette condition reste suffisante pour assurer que A est relativement compact, même si E n'est plus supposé semi-compact.*

De la démonstration du théorème 2, ou du théorème 2 directement, on déduit immédiatement le résultat suivant:

PROPOSITION 1. *Soit E un espace semi-compact, E un espace complètement régulier, A une partie relativement compacte de l'espace $C_s(E, F)$. Alors A est encore relativement compact dans l'espace $C_s(\tilde{E}, F)$, où $\tilde{E}$ est l'espace obtenu en munissant E de la topologie la moins fine rendant continues les applications éléments de A. En particulier, toute application f de E dans F que est limite simple d'applications éléments de A, est encore continue au sens de la topologie de $\tilde{E}$* (c'est à dire que, pour tout $x_0 \varepsilon E$ et tout voisinage V de $f(x_0)$ dans F, il existe un nombre fini d'éléments $f_i \varepsilon A$ et des ouverts Ω_i dans F, tels que $f_i(x_0) \varepsilon \Omega_i$ pour tout i, et que $f_i(x) \varepsilon \Omega_i$ pour tout i entraine $f(x) \varepsilon V$.[2]

Signalons pour être complet un autre cas intéressant et non classique où la semi-compacité relative entraine la compacité relative, qui m'a été signalé par M. J. Dieudonné:

THÉORÈME 3. *Soit E un espace complètement régulier dont la topologie $\mathfrak{T}$ soit plus fine qu'une certaine topologie métrisable $\mathfrak{T}_0$. Alors dans E les parties (relativement) compactes, (relativement) semi-compactes et (relativement) strictement semi-compactes sont identiques, et leur topologie métrisable.*

---

[2] En fait, cet énoncé est loin d'être profond, du moins si E est compact. On peut en effet montrer alors par voie directe le résultat bien moins restrictif: Si E est *compact*, F un espace topologique *séparé quelconque*, A un ensemble *quelconque* d'applications continues de E dans F, alors toute application *continue* de E dans F qui est limite simple d'applications éléments de A, est déjà continue quand on munit E de la topologie la moins fine rendant continues les applications éléments de A.

Il suffit de montrer que toute partie relativement semi-compacte A est relativement compacte. Car alors, son adhérence $\bar{A}$ étant compacte et la topologie induite par $\mathfrak{T}_0$ sur $\bar{A}$ étant séparée et moins fine que celle induite par $\mathfrak{T}$, elle doit lui être identique, d'où suit que $\bar{A}$ est métrisable et *strictement* semi-compact. Tout revient donc à montrer que tout ultra-filtre ϕ sur A converge vers quelque $x_0 \varepsilon E$. Mais A étant aussi relativement semi-compact pour $\mathfrak{T}_0$ qui est métrisable, A est relativement compact pour $\mathfrak{T}_0$, donc ϕ converge pour $\mathfrak{T}_0$ vers un $x_0 \varepsilon E$. Tout revient à montrer que la convergence a lieu aussi au sens de $\mathfrak{T}$, donc (cf. [2] p. 11, proposition 9) que pour toute fonction numérique continue f sur E, $f(x)$ converge vers $f(x_0)$ suivant le filtre ϕ. Soit $\mathfrak{T}_f$ la topologie la moins fine sur E rendant continues f et l'application identique de E sur E muni de $\mathfrak{T}_0$, cette topologie est métrisable, plus fine que $\mathfrak{T}_0$ et moins fine que $\mathfrak{T}$. A est donc aussi relativement semi-compact pour $\mathfrak{T}_f$, donc relativement compact pour cette topologie, ϕ tend donc vers une limite $y \varepsilon E$ au sens de $\mathfrak{T}_f$, et on a forcément $y = x_0$ puisque $\mathfrak{T}_f$ est plus fine que $\mathfrak{T}_0$. Il suit bien que $f(x)$ tend vers $f(x_0)$ suivant le filtre ϕ, CQFD.

COROLLAIRE. *Soit E un espace topologique, F un espace métrique, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Supposons qu'il existe une suite d'ensembles éléments de $\mathfrak{S}$ dont la réunion soit partout dense dans E* (*en particulier, il suffit qu'il existe dans E une suite partout dense*). *Alors dans $C_{\mathfrak{S}}(E,F)$ les parties* (*relativement*) *compactes,* (*relativement*) *semi-compactes et* (*relativement*) *strictement semi-compactes sont identiques.*

Remarquons que le théorème 3 aurait pu se démontrer aussi rapidement sans l'aide des ultra-filtres, en montrant directement que sur l'adhérence des parties relativement semi-compactes, les topologies $\mathfrak{T}$ et $\mathfrak{T}_0$ sont identiques. Mais la méthode employée montre plus généralement que si on considère un ensemble de topologies $\mathfrak{T}_i$ sur E où les parties (relativement) semi-compactes soient (relativement) compactes, et si cette famille de topologies admet un plus petit élément $\mathfrak{T}_0$ séparé, alors la borne supériere $\mathfrak{T} = \operatorname{Sup}_i \mathfrak{T}_i$ satisfait à la même hypothèse que les $\mathfrak{T}_i$.

4. Appplications aux espaces vectoriels localement convexes. Les théorèmes 1 et 2 s'appliquent aux sous-espaces fermés d'espaces $C_{\mathfrak{S}}(E,F)$. De manière générale, l'application du théorème 1 peut se présenter ainsi: On donne un ensemble B d'applications d'un ensemble E dans un espace uniforme séparé F dont les parties relativement semi-compactes soient relativement compactes, examiner s'il en est de même dans B muni d'une topologie

$\mathfrak{T}_{\mathfrak{S}}$. On pourra l'affirmer dès qu'on aura trouvé sur E, pour toute partie relativement semi-compacte A de B, une topologie rendant continues les applications éléments de A, et assez peu fine pour que toute application de E dans F dont les restrictions aux parties semi-compactes de E sont continues, et qui soit par ailleurs limite au sens de $\mathfrak{T}_{\mathfrak{S}}$ d'applications éléments de A, soit élément de B. Remarque analogue pour l'application du théorème 2, quand $\mathfrak{T}_{\mathfrak{S}}$ est la topologie de la convergence simple, mais alors on a intérêt à prendre sur E une topologie aussi fine que possible donnant encore suffisamment de parties semi-compactes pour que toute application de E dans F dont les restrictions à ces parties sont continues (et de plus limite simple d'éléments de A) soit continue. Ces deux considérations se reflètent exactement dans les deux propositions qui vont suivre.

Si E est un espace vectoriel localement convexe séparé, il peut être considéré comme l'espace des formes linéaires continues sur son dual faible E', muni d'une topologie $\mathfrak{T}_{\mathfrak{S}}$ (théorème de Mackey, cf. [9] et [5]) $\mathfrak{S}$ étant un ensemble de parties convexes et *faiblement compactes* recouvrant E'. D'autre part on peut montrer ([7]) que si E est complet, tout forme linéaire sur E' dont les restrictions aux éléments de $\mathfrak{S}$ sont continues, est faiblement continue, c'est à dire élément de E. Comme par ailleurs toute limite simple d'applications linéaires est linéaire, on obtient en premier lieu la généralisation du théorème d'Eberlein annoncée au début:

PROPOSITION 2. *Si E est un espace localement convexe séparé complet ou seulement complet pour la topologie $\tau(E,E')$ de Mackey associée,* (cf. [5] et [9]) *ses parties relativement semi-compactes et relativement compactes sont identiques* (et ceci d'ailleurs manifestement pour toute topologie localement convexe sur E donnant le même dual).

En second lieu, on a le résultat

PROPOSITION 3. a) *Sous les conditions de la proposition précédente, pour qu'une partie A de E soit faiblement relativement compacte, il faut et il suffit qu'elle soit bornée, et qu'il n'existe pas de suite (x_i) extraite de A et de suite (x'_j) extraite d'une partie faiblement compacte convexe de E', telles que $\lim\limits_i \lim\limits_j \langle x_i, x'_j\rangle$ et $\lim\limits_j \lim\limits_i \langle x_i, x'_j\rangle$ existent et soient distincts.*

b) *Plus généralement, soit (K_α) une famille de parties convexes de E', relativement faiblement compactes (et non forcément fermées), telle que la famille des adhérences faibles $\overline{K_\alpha}$ engendre algébriquement E', et que E soit complet pour la topologie de la convergence uniforme sur les $\overline{K_\alpha}$. Alors le* *critère précédent de relative faible compacité de A subsiste, si on assujettit la suite (x'_i) à être extraite de quelque K_α.*

Remarque 3. On voit facilement que les propositions précédentes valent encore si on suppose non pas E complet, mais seulement ses parties bornées et fermées complètes (il suffit de passer au complété de E pour la topologie donnée ou la topologie $\tau(E, E')$). De façon plus générale encore, les propositions 2 et 3 valent pour une partie particulière A de E, dès qu'on sait que l'enveloppe convexe fermée de A est complète (ne fût-ce que pour $\tau(E, E')$ lorsqu'il s'agit de proposition 2 ou proposition 3a)).—Comme toute partie faiblement compacte de E est forcément complète pour les topologies envisagées, il ne semble pas raisonnable d'espérer généraliser encore ces derniers résultats (mais nous approfondirons encore considérablement la proposition 2 par le théorème 7 plus bas).

Il est d'ailleurs facile de construire un espace vectoriel non complet, hyperplan fortement fermé d'un dual faible d'un espace de Banach par exemple, dans lequel il y ait des parties semi-compactes non relativement compactes. Soit en effet Ω un espace localement compact et semi-compact, mais non compact (par exemple l'espace des nombres ordinaux de seconde classe), soit $\hat{\Omega}$ l'espace compact obtenu par adjonction du "point à l'infini" ϖ. Soit E l'espace des fonctions complexes continues sur $\hat{\Omega}$, muni de la norme de la convergence uniforme, E' son dual (espace des mesures de Radon sur $\hat{\Omega}$). Si on identifie tout point de $\hat{\Omega}$ avec la mass $+1$ placée en ce point, la topologie de $\hat{\Omega}$ s'identifie à la topologie induite par la topologie faible de E'. Il est manifeste que ϖ n'appartient pas au sous-espace fortement fermé engendré par Ω (sa distance à ce dernier est égale à 1), il existe donc un hyperplan fortement fermé V de E' contenant Ω et non ϖ. Dans cet espace (muni de la topologie faible), Ω est semi-compact et non faiblement relativement compact.

La proposition 3 donne un critère pour qu'une suite de E converge faiblement; il faut et il suffit en effet qu'elle soit faiblement relativement compacte, et qu'elle converge sur une partie totale E'_1 de E' (car sur une partie faiblement compacte de E, la topologie $\sigma(E, E')$ coïncide forcément avec la topologie séparée moins fine $\sigma(E, E'_1)$). Nous ne donnons pas l'énoncé explicite, qui de toutes façons pourra beaucoup s'améliorer plus bas. Mais donnons une application immédiate de la proposition 3b), due à ce que le bidual E'' d'un espace E (cf. [5]) est engendré par les adhérences faibles des parties bornées de E:

Proposition 4. *Soit E un espace localement convexe, E' son dual fort* (cf. [5]) *supposé complet, E'' le dual de E' fort. Pour qu'une partie A de E' soit relativement compacte pour $\sigma(E', E'')$ il faut et il suffit qu'elle soit fortement bornée, et qu'il n'existe pas de suite bornée (x_i) extraite de E et de suite (x'_j) extraite de A, telles que $\lim_i \lim_j \langle x_i, x'_j\rangle$ et $\lim_j \lim_i \langle x_i, x'_j\rangle$ existent et soient distincts.*

Ici encore, il suffit de supposer seulement que les parties fermées et bornées de E' fort sont complètes. Et on a encore un critère correspondant pour qu'une suite dans E' converge pour $\sigma(E', E'')$: il faut et il suffit qu'elle soit relativement compacte pour cette topologie, et qu'elle converge sur une partie totale de E.—Noter que si on suppose A faiblement relativement compact, il est inutile de supposer E' fort complet (car l'adhérence forte de A sera déjà complète).

5. Critères de semi-compacité stricte. Soit de nouveau E un espace topologique, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Pour qu'un filtre sur une partie *relativement compacte* A de $C_{\mathfrak{S}}(E, F)$ converge, il faut et il suffit qu'il converge en chaque point d'une partie partout dense E_1 de E (puisque la topologie de la convergence simple sur E_1 est encore séparée sur $C(E, F)$, et moins fine que la topologie $\mathfrak{T}_{\mathfrak{S}}$). Si on suppose seulement A relativement semi-compact, la conclusion subsiste à condition de se borner aux filtres définis par des *suites* (f_n). En effet, cette suite ne peut avoir dans $C_{\mathfrak{S}}(E, F)$ qu'une seule valeur d'adhérence, (définie par ses valeurs sur E_1), et d'autre part on vérifie immédiatement que dans une partie relativement semi-compacte d'un espace topologique séparé C, les suites convergentes sont précisément celles qui ont un seul point adhérent.

PROPOSITION 5. *Soit E un espace topologique, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E. Supposons qu'il existe une suite (E_i) de parties de E, dont la réunion soit partout dense, et telle que dans chacun des espaces $C_{\mathfrak{S}_i}(E_i, F)$ ($\mathfrak{S}_i$ désignant la trace de $\mathfrak{S}$ sur E_i), la (semi-)compacité relative d'une partie de l'espace entraine sa stricte semi-compacité relative. Alors il en est de même dans $C_{\mathfrak{S}}(E, F)$.*

Soit en effet (f_n) une suite relativement (semi-)compacte dans $C_{\mathfrak{S}}(E, F)$. Pour tout i, la suite des restrictions des f_n à E_i forme alors une suite relativement (semi-)compacte dans $C_{\mathfrak{S}_i}(E_i, F)$, ce qui permet par hypothèse d'extraire de (f_n) une suite dont les restrictions à E_i convergent dans $C_{\mathfrak{S}_i}(E_i, F)$. Par le procédé diagonal, on peut alors extraire de (f_n) une suite telle que pour *tout* i, la suite des restrictions à E_i converge dans $C_{\mathfrak{S}_i}(E_i, F)$. Cette suite converge en particulier en chacun des points de $\bigcup_i E_i$, qui est dense dans E, d'où resulte qu'elle converge dans $C_{\mathfrak{S}}(E,F)$ en vertu de nos remarques préliminaires.

Une partie du corolllaire du théorème 3 est contenu dans la proposition précédente (savoir que dans $C_{\mathfrak{S}}(E,F)$, la semi-compacité relative entraine la stricte semi-compacité relative, sous les hypothèses spécifiées dans ce corollaire). La partie la plus profonde du corollaire en question échappe pourtant à la proposition 5, en revanche nous avons le

COROLLAIRE. *Si E contient une suite partout dense, et si dans F toute partie relativement (semi-)compacte est strictement semi-compacte, alors il en est de même dans $C_{\mathfrak{S}}(E,F)$.*

Mais on notera que quelque simple que soit l'espace F (par exemple le segment compact $(0,1)$), pour avoir des résultats dans le genre du précédent, il faut faire quelque hypothèse sur le couple $(E,\mathfrak{S})$. Ainsi, si E est un espace discret non dénombrable, et $F=(0,1)$ on sait bien que le produit topologique $C_s(E,F)$ est compact, mais non strictement semi-compact.—Il est tout aussi évident que la moindre des choses qu'il faille supposer sur F pour avoir un résultat, c'est que dans F lui-même toute partie relativement compacte soit strictement relativement semi-compacte.

Le théorème suivant tire son intérèt du fait qu'il ne fait intervenir aucune condition de dénombrabilité sur l'espace E lui-même:

THÉORÈME 4. *Soit E un espace compact, F un espace uniforme séparé, $\mathfrak{S}$ un ensemble de parties de E recouvrant E, A une partie de $C_{\mathfrak{S}}(E,F)$ relativement semi-compacte. Supposons de plus que pour toute $f\in A$, le sous-espace $f(E)$ de F ait une topologie métrisable (il suffit donc que F ait une topologie métrisable). Alors A est strictement relativement semi-compacte.*

Soit (f_n) une suite extraite de A, tout revient à montrer qu'on peut en extraire une suite qui converge en chaque point. On est donc ramené au cas de la topologie $\mathfrak{T}_s$ de la convergence simple, et nous supposerons maintenant que F est un espace topologique séparé quelconque.—Soit B l'adhérence dans $C_s(E,F)$ de l'ensemble des f_n, considérons sur E la topologie $\mathfrak{T}'$ la moins fine rendant continues les applications éléments de B, manifestement la suite (f_n) est encore relativement semi-compacte dans l'espace $C_s(\tilde{E},F)$, où $\tilde{E}$ désigne E muni de $\mathfrak{T}'$. Pour qu'une suite extraite de (f_n) converge en chaque point, il suffit donc qu'elle converge en chaque point d'une partie dense de $\tilde{E}$, et l'application du procédé diagonal nous ramène à montrer qu'il existe dans $\tilde{E}$ une suite partout dense. Mais $\phi(x)=\{f(x)\}_{f\in B}$ étant l'application canonique de E dans le produit topologique $G = \prod_{f \in B} f(E)$, il revient manifestement au même de montrer que l'image $K = \phi(\tilde{E})$ admet une suite partout dense. Mais K étant compact (comme image *continue* du *compact* E) sa topologie est aussi la moins fine de celles qui rendent continues les applications f_n (topologie qui est en effet moins fine, et d'autre part *séparée* comme on vérifie aussitôt). K s'identifie donc à un sous-espace du produit topologique $\prod_n f_n(E)$, qui est *métrisable,* par conséquent K est un compact métrisable, et à fortiori séparable.

Remarque 4. Le théorème 4 vaut encore si on suppose seulement que E est semi-compact. Tout revient en effet à montrer que K est compact, mais K est déjà semi-compact comme image continue de E, d'autre part la topologique de K est complèment régulière et plus fine que la topologie métrisable définie par les f_n; la compacité de K resulte alors du théorème 3.— On aurait aussi pu s'épargner ce raisonnement et abréger en même temps la démonstration précédente en faisant usage de la proposition 1, qui dit que la suite (f_n) est encore relativement semi-compacte dans l'espace $C(\tilde{E}, F)$, lorsque $\tilde{E}$ désigne E muni de la topologie la moins fine rendant continues les f_n; tout revient alors à trouver une suite dense dans $\tilde{E}$, ce qui est immédiat.

En conjuguant le théorème 4 et la proposition 5, on obtient des cas étendus où la semi-compacité relative entraine la semi-compacité relative stricte. Le théorème de Šmulian pour la topologie faible des espaces de Banach et plus généralement des espaces $(\mathring{\mathfrak{F}})$ (cf. [5]) en est un cas particulier, puisque un espace $(\mathfrak{F})$ s'identifie à l'espace des formes linéaires continues sur son dual faible E', et que E' est réunion d'une suite de parties faiblement compactes. On notera d'ailleurs la parenté entre la démonstration directe du théorème de Šmulian, et celle du théorème 4. Donnons pour être complet l'énoncé le plus général du théorème de Šmulian (énoncé qui peut d'ailleurs se démontrer directement comme dans le cas classique) :

PROPOSITION 6. *Soit E un espace localement convexe, (x_n) une suite faiblement relativement semi-compacte dans E, K une partie faiblement compacte du dual E'. Alors on peut extraire de (x_n) une suite qui converge en chaque point de K (et par conséquent, en chaque point du sous-espace vectoriel faiblement fermé de E' engendré par K).—Si dans E il existe une suite de voisinages de l'origine dont l'intersection soit réduite à $\{0\}$, alors on peut extraire de (x_n) une suite faiblement convergente.*

(il suffit de noter que la dernière hypothèse assure l'existence dans E'

d'une suite de parties faiblement compactes dont la réunion soit partout dense).—Rappelons que déjà dans le dual faible d'un espace de Banach peut exister une suite relativement faiblement compacte (c'est à dire bornée), dont aucune suite extraite ne converge faiblment (cf. [5]), de sorte qu'une telle situation ne peut pas être considérée comme tératologique.

6. Critères de compacité faible dans les espaces $C^\infty(E)$. Si E est un espace topologique, nous désignons par $C(E)$ l'espace des fonctions complexes continues sur E, par $C^\infty(E)$ l'espace des fonctions complexes continues et bornées sur E, muni de la norme uniforme qui en fait un espace de Banach. Si E est compact ou semi-compact, les ensembles $C(E)$ et $C^\infty(E)$ coïncident, et nous désignerons l'espace de Banach $C^\infty(E)$ par $C(E)$ pour abréger.

THÉORÈME 5. *Soit E un espace compact, pour qu'une partie A de $C(E)$ soit faiblement relativement compacte, il faut et il suffit qu'elle soit bornée, et relativement compacte dans $C(E)$ pour la topologie de la convergence simple.*

La nécessité de la condition est manifeste. Pour montrer qu'elle est suffisante, il suffit de montrer d'après le théorème d'Eberlein (cf. plus haut) que de toute suite (f_n) extraite de A on peut extraire une suite faiblement convergente. Mais comme une suite (g_n) extraite de (f_n) est uniformément bornée par hypothèse, et que par conséquent (les formes linéaires continues sur $C(E)$ n'étant autres que les mesures de Radon sur E) sa convergence faible équivaut à sa convergence en chaque point (théorème de Lebèsgue), il suffit donc d'extraire de la suite (f_n), relativement compacte pour la topologie de la convergence simple, une suite (g_n) qui converge en chaque point. Mais cela est possible en vertu du théorème 4.

Il faut bien noter que ce théorème n'est plus exact lorsqu'on substitue à la topologie de la convergence simple une topologie strictement moins fine, comme par exemple la topologie de la convergence en tout point sauf un seul x_0, comme on s'en convainc sans difficulté. Le théorème qui correspond au précédent et au suivant dans les espaces localement convexes généraux sera examiné en détail plus bas.

Le Théorème 5 permet l'application des critères de compacité établis au théorème 2, et notamment le critère d), qui ne fait intervenir que les valeurs des fonctions sur une partie dense de E. On a même le

THÉORÈME 6. *Soit E un espace topologique quelconque. Pour qu'une partie A de $C^\infty(E)$ soit relativement faiblement compacte, il faut et il suffit qu'-*

elle soit bornée, et qu'il n'existe pas de suite (x_i) *extraite de* E *de suite* (f_j) *extraite de* A *telles que* $\lim_i \lim_j f_j(x_i)$ *et* $\lim_j \lim_i f_j(x_i)$ *existent et soient distincts. Ce critère subsiste si on assujettit la suite* (x_i) *à être extraite d'une partie dense* E_1 *de* E.

On sait que l'espace $C^\infty(E)$ s'identifie à l'espace des fonctions complexes continues sur la " compactification de Čech " $\hat{E}$ de E (qui s'identifie aussi à l'espace des " caractères " de l'algèbre normée complète $C^\infty(E)$—mais en fait la théorie est très élémentaire, cf. par exemple N. Bourbaki [2], page 14, exercices 6 et 7). Il existe une application canonique continue $x \to \tilde{x}$ de E sur une partie partout dense $\tilde{E}$ de $\hat{E}$ (application qui est biunivoque si et seulement si E est complètement régulier, mais peu importe), telle que l'on ait $f(x) = \tilde{f}(\tilde{x})$ quels que soient $x \in E$ et $f \in C^\infty(E)$ (où $\tilde{f}$ est la fonction sur $\hat{E}$ définie par f). D'ailleurs, l'image $\tilde{E}_1$ de E_1 dans $\hat{E}$ sera donc aussi dense. Il suffit alors d'appliquer le théorème 5 à l'espace $C(\hat{E})$, puis le corollaire 2 du théorème 2 à ce même espace et la partie dense $\tilde{E}_1$ de $\hat{E}$.— Notons que l'application de ce dernier théorème et du critère du corollaire 2 du théorème 2, montre aussitôt que le théorème 5 reste valable si E est seulement semi-compact.

Donnons une application immédiate du théorème 6. Si G est un semi-groupe, muni éventuellement d'une topologie qui rende continues ses translations à gauche et à droite, nous dirons avec *F.* Eberlein ([6]) qu'une fonction complexe bornée et continue sur G est faiblement presque-périodique à gauche (resp. à droite), si l'ensemble de ses translatées gauches (respectivement droites) est une partie relativement faiblement compacte de l'espace de Banach $C^\infty(G)$. On a alors immédiatement la

PROPOSITION 7. *Pour qu'une* $f \in C^\infty(G)$ *soit faiblement presque-périodique à gauche (ou à droite) il faut et il suffit qu'il n'existe pas de suites* (x_i) *et* (y_j) *extraites de* G, *telles que* $\lim_i \lim_j f(x_iy_j)$ *et* $\lim_j \lim_i f(x_iy_j)$ *existent et soient distincts. En particulier, les fonctions faiblement presque périodiques à gauche et à droite sont les mêmes.* Il sera donc à propos de les appeller fonctions faiblement presque-périodiques tout court).

7. Renforcement des critères de faible compacité relative dans les espaces vectoriels localement convexes.

THÉORÈME 7. *Soit* E *un espace vectoriel localement convexe séparé,* (K_α) *une famille de parties du dual* E' *de* E, *à enveloppes convexes cerclées* *relativement faiblement compactes, et telles que la famille des enveloppes convexes cerclées fermées* $\widetilde{K}_\alpha$ *des* K_α *engendre algébriquement tout* E'. *Soit* A *une partie bornée de* E, *et supposons* E *complet et pour la topologie* $\mathfrak{T}$ *de la convergence uniforme sur les* K_α, *ou du moins l'enveloppe convexe fermée de* A *complète pour cette topologie.*

a) *Si les* K_α *sont faiblement fermés (c'est à dire faiblement compacts) alors pour que* A *soit faiblement relativement compact dans* E, *il faut et il suffit que pour tout* α, *l'ensemble des fonctions continues sur* K_α *définies par les éléments de* A *soit relativement compact dans* $C(K_\alpha)$ *pour la topologie de la convergence simple.*

b) *Si on ne suppose plus forcément les* K_α *fermés, une condition nécessaire et suffisante pour que* A *soit relativement faiblement compact, est qu'il n'existe pas de suite* (x_i) *extraite de* A *et de suite* (x'_j) *extraite de quelque* K_α, *telles que* $\lim_i \lim_j \langle x_i, x'_j \rangle$ *et* $\lim_j \lim_i \langle x_i, x'_j \rangle$ *existent et soient distincts.*

En vertu du théorème 2 corollaire 2 (qui s'applique ici puisque A est borné), la condition énoncée dans b) équivaut à la condition énoncée dans a), appliquée aux adhérences faibles des K_α, de sorte qu'on peut se borner à démontrer a). Nous identifions comme d'habitude E à l'espaces des formes linéaires continues sur son dual faible, et notons comme dans 4. que tout revient à montrer que pour toute forme linéaire X sur E' qui est faiblement adhérente à A, les restrictions aux $\widetilde{K}_\alpha$ sont faiblement continues. (Dans la suite, il est inutile de conserver l'indice α). D'après le théorème de Mackey ([9]), le dual de E muni de $\mathfrak{T}$ est encore E'. Il existe d'autre part une application linéaire canonique $x \to u(x)$ de E dans l'espace de Banach $C(K)$ des fonction complexes continues sur K, application qui est continue par la définition même de $\mathfrak{T}$, et dont la transposée u' est donc une application faiblement continue du dual C' de $C = C(K)$ dans E'. L'image de la boule unité B de C' par u' est donc une partie convexe cerclée faiblement compacte de E' (puisque B est faiblement compacte), contenant évidemment K, donc aussi $\widetilde{K}$. En fait, il nous sera commode de savoir qu'elle est même identique à $\widetilde{K}$, cela resulte immédiatement du fait connu que B est l'enveloppe convexe cerclée faiblement fermée dans C' de l'ensemble des " masses $+1$ " placées aux divers points de K (comme il resulte aussitôt de l'emploi des ensembles polaires, cf. [5]). Nous allons montrer que la restriction de X à $\widetilde{K}$ est de la forme $\langle X, u'.\mu \rangle = \langle f, \mu \rangle$, μ désignant l'élément générique de B, et où f est un élément convenable de l'espace $C = C(K)$ (c'est en fait la fonction sur $K: f(x') = \langle X, x' \rangle$), il s'ensuivra aussitôt que la restriction de X à $\widetilde{K}$ est continue, puisqu'en la composant avec l'application continue u' du compact B *sur* $\tilde{K}$, on obtient une application continue (cf. [1] page 53, th. 1, et page 62 th. 1, cor. 2). Soit donc ϕ la trace sur A du filtre des voinages faibles de X, on a pour tout $x' = u'.\mu (\mu \in B)$:

$$\langle X, u'.\mu\rangle = \lim_{\phi}.\langle x, u'.\mu\rangle = \lim_{\phi}.\langle u.x, \mu\rangle$$

or l'image de ϕ par u est un filtre de Cauchy pour la convergence simple, et $u(A)$ est *faiblement* relativement compact dans $C(K)$, comme il resulte de l'hypothèse et au théorème 5; il suit que $u.x$ tend *faiblement* suivant ϕ vers une limite $f \in C(K)$, d'où suit bien $\langle X, u'.\mu\rangle = \langle f, \mu\rangle$.

COROLLAIRE 1. *Soit E un espace de Banach, K l'ensemble des points extrémaux de la boule unité de son dual. Pour que $A \subset E$ soit faiblement relativement compact, il faut et il suffit qu'il n'existe pas de suite (x_i) extraite de A et de suite (x'_j) extraite de K, telle que* $\lim_i.\lim_j.\langle x_i, x'_j\rangle$ *et* $\lim_j \lim_i.\langle x_i, x'_j\rangle$ *existent et soient distincts.*

COROLLAIRE 2. *Soit E un espace localement convexe, (B_α) une famille de parties bornées de E telle que toute partie bornée de E soit contenue dans l'enveloppe convexe cerclée fermée de quelque B_α. Supposons le dual fort* (cf. [5]) *E' de E complet, ou du moins ses parties bornées et fermées complètes. Pour que $A \subset E'$ soit relativement compact pour la topologie $\sigma(E', E'')$ (E'' désignant le dual de E' fort) il faut et il suffit qu'elle soit fortement bornée, et qu'il n'existe pas de suite (x_i) extraite de quelques B_α et de suite (x'_j) extraite de A, telles que* $\lim_i.\lim_j.\langle x_i, x'_j\rangle$ *et* $\lim_j.\lim_i.\langle x_i, x'_j\rangle$ *existent et soient distincts.*

Remarque 5. En fait, sous les conditions du théorème 7, on peut même affirmer que l'enveloppe convexe fermée de A est faiblement compacte. En effet, la démonstration d'un théorème connu de Krein pour les espaces de Banach se transpose aux espace vectoriels localement convexes pour donner l'énoncé suivant: Soit E un espace localement convexe séparé, A une partie faiblement relativement compacte; pour que son enveloppe convexe fermée soit faiblement compacte, il faut et il suffit qu'elle soit complète (ne fût-ce d'ailleurs que pour la topologie $\tau(E, E')$ associée).—Nous ne donnerons pas la démonstration de cette proposition, qui s'appuie essentiellement sur le théorème d'Eberlein généralisé (proposition 2) et le resultat de [7] rappelé plus haut qui nous a déjà servi pour la proposition 2 et le théorème 7.

Notons encore que le théorème 6 donne comme corollaire immédiat un critère de convergence faible d'une suite dans un espace $C^\infty(E)$, et le théorème 7 un critère de convergence faible d'un suite dans un espace de Banach quelconque. Ces critères, pour le cas particulier de suites tendant faiblement vers 0, se trouvent déjà dans Banach ([4], page 222). Il ne semble pas possible d'ailleurs d'en déduire les théorèmes 6 et 7.

INSTITUT DE MATHÉMATIQUES, NANCY (M. ET M.), FRANCE.

BIBLIOGRAPHIE.

[1] N. Bourbaki, Topologie Générale, Chap. 1 et 2, *Actualités Scient. et Ind.*, No. 858 (1940).

[2] ———, Topologie Générale, Chap. 9, *ibid.*, No. 1045.

[3] ———, Topologie Générale, Chap. 10, *ibid.*, No. 1084.

[4] S. Banach, Théorie des opérations linéaires, Varsovie, 1932.

[5] J. Dieudonné et L. Schwartz, " La dualité dans les espaces $(\mathfrak{F})$ et $(\mathfrak{L}\mathfrak{F})$," *Annales de l'Institut Fourier* (Grenoble), vol. 1 (1950), pp. 61-101.

[6] F. Eberlein, "Abstract ergodic theorems and weak almost periodic functions," *Transactions of the American Mathematical Society*, vol. 67 (1949), pp. 217-240.

[7] A. Grothendieck, " Sur la complétion du dual d'un espace localement convexe," *Comptes Rendus*, vol. 230, pp. 605-606.

[8] G. Köthe, " Erweiterung von Linearfunktionen in linearen Räumen," *Mathematische Annalen*, vol. 116 (1939), pp. 719-732.

[9] G. W. Mackey, " On convex topological linear spaces," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 520-537.

[10] A. Weil, " Sur les espaces à structure uniforme et la Topologie Générale," *Actualités Scient. et Ind.*, No. 551 (1937).

GROUPS AND CARDINAL NUMBERS.*

By W. R. Scott.

1. Introduction. Let G be an infinite group. For $x \varepsilon G$, let $E(x, G) = E(x)$ be the set of $g \varepsilon G$ such that the equation $g^n = x$ has no solution for n. Let $K = K(G)$ be the set of $k \varepsilon G$ such that $o(E(k)) < o(G)$, where $o(S)$ is the cardinal number of elements in S. Observe that if G is a p^∞ group, then $K(G) = G$.

This paper is devoted primarily to the problem of determining K, given G. The problem is solved for Abelian groups (Corollary 1 to Theorem 8, and Theorems 9 and 10), and considerable progress is made in the general case (Theorems 1-8, Corollary 2 to Theorem 8, and the Corollary to Theorem 11).

In the process of (partial) solution of the above problem, two sidepaths are investigated. In section **3**, a few results are given concerning the size of layers of a group. These results are perhaps of independent interest. In sections **4**, **5**, and **6**, in conjunction with the study of K, the intersection D of all subgroups G_α of G with $o(G_\alpha) = o(G)$ is introduced. The principal result is that $D = K$ for Abelian groups G.

2. Definitions and notations. Let G be a group written multiplicatively with identity e. If $g \varepsilon G$, $o(g)$ will denote the order of g; $o(g) = \infty$ means that $g^n = e$ implies $n = 0$. G is *periodic* if $o(g) < \infty$ for every $g \varepsilon G$. G is *locally cyclic* if for every $g_1, g_2 \varepsilon G$ there exists a $g \varepsilon G$ and integers m and n such that $g^m = g_1$, $g^n = g_2$. A subgroup H of G is *central* if $H \subseteq Z$ where Z is the center of G. A subgroup H is *strictly characteristic* if $\sigma(H) \subseteq H$ for every endomorphism $\sigma(G) = G$ of G onto G. It is *fully characteristic* if $\sigma(H) \subseteq H$ for every endomorphism σ of G. The letter p will always denote a prime. A *p-group* is a group G such that $o(g) = p^{r(g)}$ for all $g \varepsilon G$, for a fixed prime p. A *layer* $L(n)$ $(L(\infty))$ of G is the set of $g \varepsilon G$ with $o(g) = n(\infty)$. The notations $o(S)$, $E(x)$, K, and $D(G) = D$ will be used as in section **1**. If G is Abelian, the *torsion* T of G is the subgroup of $g \varepsilon G$ with $o(g) < \infty$. The *p-component* T_p of T is defined by the equation $T_p = \bigcup_{i=0}^{\infty} L(p^i)$. For Abelian groups G, let $H(p, r) = \bigcup_{i=0}^{r} L(p^i)$. A p^∞

* Received January 15, 1951; revised October 9, 1951.

group is a p-component of the group of rationals mod 1. A direct product of subgroups of a group will be denoted by Π. *The symbol* $\bigcup$ *will always denote the point set union.* If S_1 and S_2 are sets, $S_1 - S_2$ will denote the set of s such that $s \varepsilon S_1$, $s \not\varepsilon S_2$. For the sake of brevity, a cardinal A will be called *standard* if (i) $A > \aleph_0$ and (ii) A is not the sum of $\aleph_0$ smaller cardinals. In the statement of a theorem, **H.** and **C.** will mean hypothesis and conclusion respectively.

3. The layers. LEMMA 1. **H.** *G is an infinite group, P a set of primes.* $S = (\bigcup_{\lambda} \bigcup_{p \not\varepsilon P} L(\lambda p)) \cup L(\infty)$.

C. $o(S) = o(G)$ *or* 0.

Proof. Suppose that there exists an $s \varepsilon S$. Clearly, we may suppose that either $o(s) = \infty$ or $o(s) = p \not\varepsilon P$. Let N be the normalizer of s. Then $i(N) = o(Cl(s))$, where $i(N)$ denotes the index of N and $Cl(s)$ the class of conjugates of s. If $g \varepsilon Cl(s)$, then $o(g) = o(s)$, and therefore $Cl(s) \subseteq S$. Hence, if $i(N) = o(G)$, we have $o(S) = o(G)$. Therefore suppose $i(N) < o(G)$. Since $o(G) = o(N)i(N)$, and since $o(G) \geqq \aleph_0$, we have $o(N) = o(G)$. If $o(N \cap S) = o(G)$, we are done. If not, then $o(N \cap (G - S) = o(G)$. Let $g \varepsilon N \cap (G - S)$. Then $(gs)^r = e$, $r \neq 0$, implies $g^r s^r = e$, which implies that s^r is in the subgroup $\{g\}$ generated by g. If $o(s) = \infty$. this would imply that $o(g) = \infty$, a contradiction since $g \not\varepsilon S$. If $o(s) = p \not\varepsilon P$, then $p | r$, whence $p | o(gs)$. Thus $gs \varepsilon S$ for all $g \varepsilon N \cap (G - S)$. Hence

$$o(S) \geqq o(N \cap (G - S)) = o(G),$$

i. e. $o(S) = o(G)$, and the theorem is proved.

COROLLARY 1. **H.** *G is a non-periodic group.*

C. $o(L(\infty)) = o(G)$.

Proof. Since G is non-periodic it is infinite. In Lemma 1, let P be the set of all primes.

COROLLARY 2. **H.** *p is a fixed prime. G is an infinite periodic group which is not a p-group.* $S = \bigcup_{n \neq p^r} L(n)$.

C. $o(S) = o(G)$.

LEMMA 2. **H.** $n = \prod_{i=1}^{r} p_i^{\alpha_i}$, *with $p_i \neq p_j$ for $i \neq j$, and $\alpha_i > 0$. G is a group (finite or infinite).*

C. $o(L(n)) \leqq \prod_{i=1}^{r} o(L(p_i^{\alpha_i}))$.

Proof. If $g \in L(n)$, then $g = \prod_{i=1}^{r} g_i$ where $g_i \in L(p_i^{\alpha_i})$.

LEMMA 3. **H.** *G is an Abelian p-group.*

C. $o(L(p^r)) \leqq o(L(p^{r-1}))(o(L(p))+1) \leqq (o(L(p))+1)^r$, $r = 1, 2, \cdots$.

Proof. If $g_1{}^p = g_2{}^p$, then $(g_1 g_2{}^{-1})^p = e$, and conversely. Hence, if $g \in G$, there are 0 or $(o(L(p))+1)$ solutions of the equation $x^p = g$. Therefore there are at most $o(L(p^{r-1}))(o(L(p))+1)$ elements $x \in G$ such that $x^p \in L(p^{r-1})$; i. e. $o(L(p^r)) \leqq o(L(p^{r-1}))(o(L(p))+1)$. The second inequality follows from the first by induction.

COROLLARY 1. **H.** *G is an Abelian p-group. $o(G) > \aleph_0$. $H(p,1) = L(1) \cup L(p)$.*

C. $o(H) = o(G)$.

Proof. If H is finite, then by Lemma 3, G is denumerable, a contradiction. Hence, again by Lemma 3,

$$o(G) = \sum_{i=0}^{\infty} o(L(p^i)) = o(L(p)) = o(H).$$

COROLLARY 2. **H.** *G is Abelian and periodic; $o(G)$ is standard.*

C. *There exist a p such that the subgroup $H(p,1)$ satisfies the relation*

$$o(H(p,1)) = o(G).$$

Proof. $G = \Pi G_p$ where G_p is the p-component of G. Hence

$$o(G) \leqq \sum_{r=1}^{\infty} \prod_{i=1}^{r} o(G_{p_i}).$$

Since $o(G)$ is standard, this implies that $o(G_p) = o(G) > \aleph_0$ for some p. The theorem follows from Corollary 1.

Remark. The condition that $o(G)$ be standard cannot be omitted (for infinite groups) in this corollary.

4. Formulas and lemmas. We first list a few elementary identities and inclusion relations for $E(x)$.

(1) $E(e)$ is the empty set.

(2) $E(x^{-1}) = E(x)$.

For $g^n = x$ if and only if $g^{-n} = x^{-1}$.

(3) $E(x_1x_2) \subseteq E(x_1) \cup E(x_2)$.

For $g^m = x_1$, $g^n = x_2$ imply $g^{m+n} = x_1x_2$.

(4) $E(\sigma(x), \sigma(G)) \subseteq \sigma(E(x, G))$ for any endomorphism σ of G.

For $g^n = x$ implies $(\sigma(g))^n = \sigma(x)$.

(5) $E(h, H) = H \cap E(h, G)$ if $h \varepsilon H \subset G$.

(6) $E(x) \supseteq G - N(x)$, where $N(x)$ is the normalizer of x.

For if $g^n = x$, then $gx = xg$, and $g \varepsilon N(x)$.

THEOREM 1. **H.** *G is an infinite group.*

C. *D is a strictly characteristic subgroup of G. Moreover $K \subseteq D$.*

Proof. Let $d \varepsilon D$, let $\sigma(G) = G$ be an endomorphism of G onto G, and let G_α be a subgroup of G with $o(G_\alpha) = o(G)$. Then if $H = \sigma^{-1}(G_\alpha)$, H is a subgroup of G with $o(H) = o(G)$. Thus $d \varepsilon H$, and therefore $\sigma(d) \varepsilon G_\alpha$. Therefore $\sigma(d) \varepsilon \cap G_\alpha = D$. Hence D is strictly characteristic.

If there exists a $k \varepsilon K - D$, then $k \not\varepsilon G_\alpha$ for some subgroup G_α with $o(G_\alpha) = o(G)$. It follows easily that $G_\alpha \subseteq E(k)$. This gives $o(E(k)) = o(G)$ which is a contradiction. Thus $K \subseteq D$.

COROLLARY. **H.** *G is an infinite group such that $G = \Pi G_\alpha$ with $o(G_\alpha) < o(G)$.*

C. *$K = D = e$.*

Proof. Let $G_\alpha{}^* = \prod_{\beta \neq \alpha} G_\beta$. Then $G = G_\alpha \times G_\alpha{}^*$, and therefore $o(G_\alpha{}^*) = o(G)$. Hence by Theorem 1, $K \subseteq D \subseteq \cap G_\alpha{}^* = e$.

LEMMA 4. **H.** *G_n is an infinite group, $n = 1, 2, \cdots$; $G_1 \subseteq G_2 \subseteq \cdots$; $G = \cup G_n$.*

C. *$K(G) \subseteq \text{Lim inf } K(G_n)$.*

Proof. Let $k \varepsilon K(G)$. Then $o(E(k)) < o(G) = \lim o(G_n)$. Hence there exists an $n_0(k)$ such that if $n > n_0$, then (i) $k \varepsilon G_n$, (ii) $o(E(k, G)) < o(G_n)$. Hence by (5), $o(E(k, G_n)) \leqq o(E(k, G)) < o(G_n)$ and $k \varepsilon K(G_n)$ if $n > n_0$, i. e. $k \varepsilon \lim\inf K(G_n)$.

Remark. If $o(G)$ is standard, then $K(G) = \lim K(G_n)$ and $D(G) = \lim D(G_n)$.

LEMMA 5. **H.** *H is a subgroup of G.* $o(G) = o(H) \geqq \aleph_0$.

C. $K(G) \subseteq K(H)$,

$D(G) \subseteq D(H)$.

Proof. By Theorem 1, $K(G) \subseteq H$. By (5), if $k \varepsilon K(G)$, then $o(E(k, H)) \leqq o(E(k, G)) < o(G) = o(H)$. Hence $k \varepsilon K(H)$. The other half of the theorem follows from the definition of D.

LEMMA 6. **H.** $o(G) \geqq \aleph_0$. *H is a proper subgroup of G.*

C. $o(G - H) = o(G)$.

Proof. If $o(H) < o(G)$, this is immediate. If $o(H) = o(G)$. then for any $g \varepsilon G - H$, $Hg \subseteq G - H$ and we have

$$o(G) \geqq o(G - H) \geqq o(Hg) = o(H) = o(G).$$

5. The subgroup K, general case.

THEOREM 2. **H.** *G is an infinite group.*

C. *K is a central, locally cyclic, periodic subgroup of G.*

Proof. (i) K is a subgroup of G. By (1), $e \varepsilon K$. By (2), if $k \varepsilon K$, then $k^{-1} \varepsilon K$. By (3), if $k_1, k_2 \varepsilon K$, then $o(E(k_1 k_2)) \leqq o(E(k_1)) + o(E(k_2)) < o(G)$, and $k_1 k_2 \varepsilon K$.

(ii) K is central. For let $k \varepsilon K$, and let N be the normalizer of k in G. Then by (6) and Lemma 6, if $N \neq G$, we get $o(E(k)) \geqq o(G - N) = o(G)$, which is a contradiction. Hence $N = G$, i. e. $k \varepsilon Z$. Thus $K \subseteq Z$ and K is central.

(iii) K is locally cyclic. Let $k_1, k_2 \varepsilon K$, $k_i \neq e$ (the other case is trivial). Then there exists an $x \varepsilon G - (E(k_1) \cup E(k_2))$. Hence there exists integers n_1, n_2 such that $x^{n_1} = k_1$, $x^{n_2} = k_2$. Let $d = (n_1, n_2)$, and let $d = r_1 n_1 + r_2 n_2$. Hence $x^d = k_1{}^{r_1} k_2{}^{r_2} \varepsilon K$. Moreover $k_i = (x^d)^{n_i/d}$. Thus K is locally cyclic.

(iv) K is periodic. Suppose, on the contrary, that there exists a $k \varepsilon K$ with $o(k) = \infty$. Then if $x \not\varepsilon E(k)$, there exists an $n(x)$ such that $x^{n(x)} = k$. Thus $o(x) = \infty$ also. Let S^+ be the set of $x \not\varepsilon E(k)$ for which (the unique) $n(x) > 1$, and S^- the set of $x \not\varepsilon E(k)$ for which $n(x) < -1$. Thus

$$G = E(k) \cup k \cup k^{-1} \cup S^+ \cup S^-,$$

and therefore either $o(S^+) = o(G)$ or $o(S^-) = o(G)$, or both hold. If $x \varepsilon S^+$, and r is any integer, then

$$(kx)^r = x^{r(n(x)+1)} \neq x^{n(x)} = k.$$

Thus $kx \varepsilon E(k)$, and therefore $kS^+ \subseteq E(k)$. Similarly $k^{-1}S^- \subseteq E(k)$. Since $o(kS^+) = o(S^+)$, $o(k^{-1}S^-) = o(S^-)$, this leads to a contradiction.

THEOREM 3. **H.** *G is not periodic.*

C. $K = e$.

Proof. By Corollary 1 of Lemma 1, $o(L(\infty)) = o(G)$. If $k \varepsilon K$, $k \neq e$, then $o(k) < \infty$ by Theorem 2. Hence $L(\infty) \subseteq E(k)$, and $o(E(k)) = o(G)$, a contradiction. Therefore $K = e$.

THEOREM 4. **H.** *$o(G)$ is standard.*

C. *K is cyclic of order p^n for some prime p and some integers n.*

Proof. By Theorem 3, we may assume that G is periodic. Therefore $o(G) = \Sigma o(L(n))$. Since $o(G)$ is standard, $o(L(n)) = o(G) > \aleph_0$ for some integer n. Therefore by Lemma 2, $o(L(p^r)) = o(G)$ for some prime p and some positive integer r. If $x \notin \bigcup_{i=0}^{r} L(p^i)$, then $L(p^r) \subseteq E(x)$ and $x \notin K$. Thus $K \subseteq \bigcup_{i=0}^{r} L(p^i)$ (actually $K \subseteq \bigcup_{i=0}^{r-1} L(p^i)$). Now since K is locally cyclic and periodic, it is isomorphic to a subgroup of the group R_1 of rationals mod 1. Hence K is cyclic of order p^n, $0 \leqq n \leqq r$ (actually $n < r$).

THEOREM 5. **H.** *G is an infinite group.*

C. *K is either a cyclic group of order p^n or a p^∞ group.*

Proof. Case 1. $o(G) > \aleph_0$. If $o(G)$ is standard, the conclusion is immediate from Theorem 4. If $o(G)$ is not standard, then there exists an increasing sequence $\{B_n\}$ of standard cardinals whose sum is $o(G)$. Choose a sequence $\{G_n\}$ of subgroups such that (α) $G_n \subseteq G_{n+1}$, (β) $o(G_n) = B_n$, and (γ) $G = \cup G_n$. Then by Theorem 4, $K(G_n)$ is cyclic of order $p_n^{t_n}$. It follows readily that $\liminf K(G_n)$ is a subgroup of G and is one of the two types described in the theorem. By Lemma 4, $K(G)$ is also of the required type.

Case 2. $o(G) = \aleph_0$. By Theorem 3, we may assume that G is periodic. If $o(L(p)) > 0$ for an infinity of primes p, then for any $x \neq e$ we have $o(E(x)) = \aleph_0$, and therefore $K = e$. If for some p, $S_p = \bigcup_{r=0}^{\infty} L(p^r)$ is infinite, then it easily follows that K is a p-group. But a locally cyclic p-group is one of the two types described. If, finally, S_p is finite for all p, and $o(S_p) = 1$ for all but a finite number of p, say $p_1, \cdots, p_r$, then there is a maximum β_i

such that there exist $g \varepsilon S_{p_i}$ with $o(g) = p_i^{\beta_i}$. Then no element of G has order greater than $n = \Pi p_i^{\beta_i}$. Hence $o(G) = \sum_{j=1}^{n} o(L(j))$ which is finite by Lemma 2. This is a contradiction. Hence the theorem is true in any case.

Note that Theorem 5 improves Theorem 2.

THEOREM 6. **H.** *G is an infinite group.*

C. *K is a fully characteristic subgroup of G.*

Proof. Let σ be any endomorphism of G and let N be the kernel of σ. If $o(N) = o(G)$, then $K \subseteq D \subseteq N$, and $\sigma(K) = e \varepsilon K$. If $o(N) < o(G)$, then $\sigma(G)$ is a subgroup of G with $o(\sigma(G)) = o(G)$. If $k \varepsilon K$, then by (4)

$$o(E(\sigma(k), \sigma(G))) \leqq o(\sigma(E(k, G))) \leqq o(E(k, G))$$
$$< o(G) = o(\sigma(G)).$$

Hence $\sigma(K) \subseteq K(\sigma(G))$. Now by Lemma 5, $K(G) \subseteq K(\sigma(G))$, and by Theorem 5, $K(\sigma(G))$ is either a cyclic group of order p^n or a p^∞ group. Since $o(\sigma(K)) \leqq o(K)$, and since both K and $\sigma(K)$ are subgroups of $K(\sigma(G))$, it follows that $\sigma(K) \subseteq K$. Hence K is fully characteristic.

THEOREM 7. **H.** *G is an infinite group. $G = H \times F$ where H is a p-group and F is a periodic subgroup of G such that* (i) $o(F) < o(G)$ *and* (ii) *$f \varepsilon F$ implies $p \nmid o(f)$.*

C. $K(G) = K(H)$.

Proof. By (i) $o(H) = o(G)$, whence by Lemma 5, $K(G) \subseteq K(H)$. Let $k \varepsilon K(H)$. Then if $h \varepsilon H - E(k, H)$, we have $h^n = k$ for some n. Let $o(h) = p^s$ and let $f \varepsilon F$, $o(f) = r$. Then there exists an integer λ such that $\lambda r \equiv n \pmod{p^s}$. Thus $(hf)^{\lambda r} = h^{\lambda r} = h^n = k$. Therefore $hf \varepsilon G - E(k, G)$. Hence

$$o(E(k, G)) \leqq o(E(k, H))o(F) < o(G).$$

Thus $k \varepsilon K(G)$. This proves that $K(G) = K(H)$.

THEOREM 8. **C.** *K is a p^∞ group if and only if there exists a central p^∞ subgroup C such that G/C is finite. If such a C exists, then $C = K = D$.*

Proof. Suppose first that such a C exists. Then G is periodic. Let $g_1, \cdots, g_n$ be representatives of the cosets of C. Then there exists an $r > 0$ such that $g_i^r = e$, $i = 1, \cdots, n$. Choose t such that $p^t \geqq r$. Then if $c \varepsilon C$, $o(c) = p^s$, and $c' \varepsilon C$, $o(c') \geqq p^{s+t}$, we have $(c'g_i)^r = c'^r$ which has order

$\geqq p^s$. Hence a suitable power of $c'g_i$ equals c. Thus $E(c)$ is finite, and $c \varepsilon K$. Hence $C \subseteq K$, and by Theorem 5, $C = K$. Also, clearly $K = D$.

Conversely suppose that K is a p^∞ group. K is central by Theorem 2. If G/K is finite, we are done. Suppose that G/K is infinite. Then $o(G/K) = o(G)$. Now let $g \varepsilon G$, and let n be the smallest positive integer such that $g^n \varepsilon K$. Then $g^n = k$ and $k'^n = k^{-1}$ for some k, $k' \varepsilon K$. Thus $(gk')^n = e$ while $(gk')^r \not\varepsilon K$ for $0 < r < n$. Therefore if $k'' \varepsilon K$, $k'' \neq e$, then $gk' \varepsilon E(k'')$, i.e. $E(k'')$ contains at least one element of each coset of K. Thus $o(E(k'')) \geqq o(G/K) = o(G)$, a contradiction. Therefore G/K is finite as asserted.

Corollary 1. **H.** *G is abelian.*

C. *K is a p^∞ group if and only if $G = H \times F$ where H is a p^∞ group and F is finite.*

Proof. If K is a p^∞ group then (see [1], p. 767) $G = K \times F$, and by Theorem 8, F is finite. The converse is obvious.

Corollary 2. **H.** $o(G) > \aleph_0$.

C. *K is a cyclic group of order p^n.*

Proof. This follows from Theorems 5 and 8.

Corollary 3. **H.** $K_2 = K(G/K(G))$.

C. *K_2 has order 1 or is not defined.*

Proof. If K is a p^∞ group, then G/K is finite and K_2 is not defined. Otherwise K is finite. Let $Kx \varepsilon K_2$. Then if E is the set of $g \varepsilon G$ such that $Kg \varepsilon E(Kx)$, we have

$$o(E) = o(K)o(E(Kx)) < o(G).$$

If $g \not\varepsilon E \cup (\cup E(k))$, then $g^n = kx$ for some integer n and some $k \varepsilon K$. and $g^m = k^{-1}$ for some m. Hence $g^{m+n} = x$, and $g \not\varepsilon E(x)$. Therefore

$$o(E(x)) \leqq o(E) + \sum_{k \varepsilon K} o(E(k)) < o(G)$$

since K is finite. Hence $x \varepsilon K$ and K_2 has order 1.

6. Abelian groups.

Lemma 7. **H.** *$G \supseteq \Pi H_\alpha$, $\alpha \varepsilon S$, where $o(G) = o(S) \geqq \aleph_0$. R is the set of subgroups G_β of G with $o(G_\beta) = o(G)$.*

C. (i) $D = e$. (ii) $o(R) = 2^{o(G)}$.

Proof. There are $2^{o(G)}$ subset of G, hence $o(R) \leqq 2^{o(G)}$. There are $2^{o(G)}$ subsets S' of S of order $o(G)$. Then if $N(S') = \Pi H_\alpha$, $\alpha \varepsilon S'$, we have $o(N(S')) = o(G)$, and $S' \neq S''$ implies $N(S') \neq N(S'')$. Hence $o(R) = 2^{o(G)}$. Clearly, also, $D \subseteq \cap N(S') = e$.

THEOREM 9. **H.** *$o(G) > \aleph_0$, and G is Abelian. R is the set of subgroups G_α of G with $o(G_\alpha) = o(G)$.*

C. (i) $o(R) = 2^{o(G)}$. (ii) $K = D = e$.

Proof. Let T be the torsion of G.

Case 1. $o(T) < o(G)$. Then $o(L(\infty)) = o(G)$. Let $B = \{b_\alpha\}$ be a maximal set of independent elements of $L(\infty)$. We assert that $o(B) = o(L(\infty))$. For suppose that $o(B) < o(L(\infty))$ and let H be the subgroup of G generated by the b_α. Then $o(H) = o(B)$ if $o(B) \geqq \aleph_0$, and $o(H) = \aleph_0$ if $o(B) < \aleph_0$. Hence, in any case, $o(H) < o(L(\infty))$. For fixed n, if $x^n = y^n$, then $(xy^{-1})^n = e$ and $xy^{-1} \varepsilon T$. Thus there are at most $o(T)$ solutions of $x^n = h$ for fixed n and fixed $h \varepsilon H$. Hence the number of solutions of $x^n \varepsilon H$ for x, allowing n to vary, $n \neq 0$, is at most $\aleph_0 o(T) o(H) < o(G) = o(L(\infty))$. Therefore there exists an $x \varepsilon L(\infty)$-H such that the set $B' = B \cup x$ is independent. This contradicts the maximality of B. Hence $o(B) = o(L(\infty))$ as asserted. The theorem follows in this case from Lemma 7.

Case 2. $o(T) = o(G)$. Then $T = \Pi T_p$.

Case 2. 1. $o(T_p) = o(T)$ *for some* p. Then by Corollary 1 of Lemma 3, we have $o(H(p, 1)) = o(G)$. Now $H(p, 1) = \Pi C_\alpha$ where the C_α are cyclic of order p. There are clearly $o(G)$ factors C_α. Therefore by Lemma 7 the theorem is true.

Case 2. 2. $o(T_p) < o(T)$ for all p. Let $U = \Pi T_{p_i}$ for all primes p_i such that $o(T_{p_i}) > \aleph_0$. Then $o(U) = o(T) = \Sigma o(T_{p_i})$. By Case 2. 1 each T_{p_i} has $2^{o(T_{p_i})}$ subgroups $H(i)$ of order $o(T_{p_i})$. For each i, choose an $H(i) \subseteq T_{p_i}$, with $o(H(i)) = o(T_{p_i})$. Then $V = \Pi H(i)$ is a subgroup of U such that $o(V) = o(U)$. The number of subgroups formed in this manner is clearly

$$\Pi 2^{o(T_{p_i})} = 2^{\Sigma o(T_{p_i})} = 2^{o(U)} = 2^{o(G)}.$$

Moreover, it is clear from Case 2. 1 that the intersection of all the subgroups V is e.

Remark 1. This result may be combined with that in [3] to get the following theorem:

If G is an Abelian group such that for any proper subgroup H it is true that $o(H) < o(G)$, then G is either a finite group or a p^∞ group.

Remark 2. The proof of Theorem 9 may be altered slightly to prove the following generalization:

If G is a non-denumerable Abelian group, $o(G) \geqq A \geqq \aleph_0$, and $R(A)$ is the set of subgroups G_α of G with $o(G_\alpha) = A$, then $o(R) = (o(G))^A$.

THEOREM 10. **H.** *G is Abelian, $o(G) = \aleph_0$, and G cannot be expressed in the form $H \times F$ with H a p^∞ group and F finite.*

C. $D = K = e$.

Proof. Case 1. G is not periodic, i. e., there exists a $g \varepsilon G$ with $o(g) = \infty$. If $g' \varepsilon G$, $1 < o(g') < \infty$, then g' is not in the subgroup generated by g. If $g'' \varepsilon G$, $o(g'') = \infty$, then g'' is not in the subgroup generated by g''^2. Hence $D = e$.

Case 2. G is periodic. Then $G = \Pi G_p$.

Case 2. 1. $o(G_p) = o(G)$ for some p.

Case 2. 1. 1. $o(H(p,r)) = o(G)$ for some r. Then $H(p,r) = \prod_{n=1}^{\infty} C_n$ where C_n is a cyclic group. Then by Lemma 7, $D = e$.

Case 2. 1. 2. $o(H(p,r)) < o(G)$ for all r, i. e., $H(p,r)$ is finite for all r. Then (see [2], p. 102) there exists a p^∞ subgroup H of G. Hence (see [1], p. 767) $G = H \times M$. By hypothesis M is infinite, and therefore $D \subseteq H \cap M = e$.

Case 2. 2. $o(G_p) < o(G)$ for all p. Then by the Corollary to Theorem 1, $D = e$.

Thus in all cases $D = e$, and therefore $K = e$ also.

7. Miscellaneous.

THEOREM 11. **H.** *N is a normal subgroup of G, $o(G/N) \geqq \aleph_0$, $D(G/N) = eN$.*

C. $D(G) \subseteq N$.

Proof. If $o(N) = o(G)$, the conclusion follows from the definition of

$D(G)$. Suppose $o(N) < o(G)$, and let $d \varepsilon D(G)$. If H^* is a subgroup of G/N not containing dN, then its inverse image H is a subgroup of G not containing d. Hence $o(G/N) = o(G) > o(H) \geqq o(H^*)$. Thus $dN \varepsilon D(G/N)$, i. e. $dN = eN$. Hence $d \varepsilon N$.

Remark. The above theorem remains true if the letter D is replaced by K throughout.

COROLLARY. **H.** $o(G) > \aleph_0$. *Q is the commutator subgroup of G.*

C. *$D \subseteq Q$, hence $K \subseteq Q$.*

Proof. If $o(Q) = o(G)$ then $D \subseteq Q$. If $o(Q) < o(G)$, then $o(G/Q) = o(G) > \aleph_0$, and G/Q is Abelian. By Theorem 9, $D(G/Q) = eQ$, and by Theorem 11, $D(G) \subseteq Q$ as asserted.

Definition. Let G be an infinite group, and let $\aleph_0 \leqq A \leqq o(G)$. Let $K(A, G)$ be the set of $k \varepsilon G$ such that $o(E(k)) < A$. Let $D(A, G)$ be the intersection of all subgroups G_α of G such that $o(G_\alpha) \geqq A$.

Note that $K(G) = K(o(G), G)$, $D(G) = D(o(G), G)$. If $A < B$, then $K(A, G)$ is a subgroup of $K(B, G)$ and $D(A, G)$ is a subgroup of $D(B, G)$. Moreover $K(A, G) \subseteq D(A, G)$, and $K(A, G)$ is fully characteristic and $D(A, G)$ is strictly characteristic in G. It follows from Corollary 2 to Theorem 8 that if $o(G) > A \geqq \aleph_0$, then $K(A, G)$ is a cyclic group of order p^n.

UNIVERSITY OF KANSAS.

REFERENCES.

[1] R. Baer, "The subgroup of the elements of finite order of an Abelian group," *Annals of Mathematics*, vol. 37 (1936), pp. 766-781.

[2] S. N. Černikov, "Infinite groups with finite layers," *Matematičeskiĭ Sbornik*, vol. 22 (1948), pp. 101-133 (Russian).

[3] I. Szélpál, "Die unendlichen Abelschen Gruppen mit lauter endlichen echten Untergruppen," *Publicationes Mathematicae (Debrecen)*, vol. 1 (1949), pp. 63-64.

ON ISOMETRIC SURFACES.*

By Aurel Wintner.

1. The starting point of the following considerations was the observation that, in the differential geometry of surfaces, the fundamental notion of *isometry* is frequently used in a loose and misleading sense. Even when writers are careful enough to specify the assumptions of smoothness (the actual degree of differentiability, possibly the analyticity) of the two surfaces concerned, usually no mention is made of the degree of smoothness required of the mapping of the two surfaces which realizes their isometry in question.

In order to clarify this objection, a few definitions will be needed. First, if S denotes a sufficiently small (open) piece of a surface in the X-space, where $X = (x, y, z)$, and, if on a sufficiently small (u^1, u^2)-domain of a parameter plane,

$$S: \quad X = X(u^1, u^2) \tag{1}$$

is a parametrization in which the vector function $X(u^1, u^2)$ is of class C^1 and such that the vector product of $X_1 = \partial X/\partial u^1$ and $X_2 = \partial X/\partial u^2$ does not vanish, then (1) is called a C^1-parametrization of S. By a C^n-parametrization is meant[1] a C^1-parametrization in which the function $X(u^1, u^2)$ is of class C^n. If S, when given as a set of points in the X-space, has *some* C^n-parametrization, then S will be called of class C^n.

In Section 4, a corresponding manner of speaking will be used with reference to the surface S and to the function class

$$C^n(\lambda), \qquad (0 < \lambda < 1), \tag{2}$$

where λ denotes a (locally uniform) Hölder index for the n-th derivatives of $X(u^1, u^2)$, for a fixed n. Similarly, (1) will be called an analytic parametrization if it is a C^1-parametrization having the property that the function $X(u^1, u^2)$ can be represented (locally) as a convergent power series in (u^1, u^2), while S will be called analytic if it has *some* analytic parametrization. Note that every analytic S has C^n-parametrizations which are not C^{n+1}-parametrizations, where n can be chosen arbitrarily.

* Received January 16, 1951; revised October 15, 1951.

[1] If $n = 1, 2, \cdots$. The limiting case $n = 0$, which (unless the contrary is implied) will be excluded, refers to one-to-one continuous parametrizations (1) of S.

Besides S, consider another S, say S'. Suppose that both S and S' are of class C^1, at least. If they are of class C^n, where $n = \infty$ is not excluded, let

$$(3) \qquad S: X(u^1, u^2) \quad \text{and} \quad S': X'(u'^1, u'^2)$$

be C^n-parametrizations; if S and S' are analytic, let their parametrizations (3) be chosen analytic. With the understanding that S and S' are sufficiently small, let

$$(4) \qquad u'^1 = u'^1(u^1, u^2), \qquad u'^2 = u'^2(u^1, u^2)$$

be a C^1-transformation (by this is meant that the two functions (4) are of class C^1 *and* have a non-vanishing Jacobian). If such a one-to-one transformation can be chosen so that it will transform the metric form, ds^2, on S into the metric form, ds'^2, on S', i. e., if the two vector functions (3) satisfy, for an appropriate choice of the mapping (4), the identity

$$(5) \qquad |\, dX'(u'^1, u'^2) \,| \equiv |\, dX(u^1, u^2) \,| \text{ by virtue of (4),}$$

then S and S' are called *isometric*. This is the definition which tacitly underlies the classical literature of the subject.

2. Needless to say, the C^1-character of the functions occurring in (3) and (4) makes (5) a meaningful statement, since *both* X and X' are functions of class C^1 in terms of (u^1, u^2) (or, equivalently, (u'^1, u'^2)) and possess therefore [2] the complete differentials [3] occurring in (5). On the other hand, even if the surfaces (3) are very smooth (say analytic), there is no justification for restricting (4) in (5) to transformations having a high degree of smoothness since nothing like such a restriction (say analyticity) is involved in the geometrical idea of a transformation (4) which preserves the metric, ds^2. This contrast leads, however, to geometrically undesirable situations.

In order to see this, consider the wording of the following assertion (stated, to be sure, because of its instructive nature only): "Two closed, convex, *analytic* surfaces, F and F', must be congruent [4] whenever they are (locally) isometric." What should be meant here by the assumption of the isometry of F and F'? The existence of

---

[2] This conclusion shows that, as far as (5) is concerned, the C^1-character of the admitted mappings (4) could slightly be generalized.

[3] Actually, not even this, the existence of the "vectors" dX, but only the existence of the "distances" $|\, dX \,|$, is needed for the formulation of the requirement (5) (so that isometry can be defined in a manner which is even more general than, but geometrically just as meaningful as, the definition referred to in the preceding footnote).

[4] Under the group consisting of the translations, rotations *and* reflections of the X-space.

(a) analytic transformations (4) or (b) just C^1-transformations (4)

which preserve the ds^2? If the statement is meant in its interpretation (a), then the content of the statement is hardly geometrical in nature, since it does not preclude the following possibility: The geometrical objects, F and F', need not be congruent if they are isometric under mappings (4) which are non-analytic but very smooth, say of class C^∞. Hence, in order to make the statement geometrically significant, its truth in interpretation (b) must be proved to imply its truth in interpretation (a).

It so happens that, in the theorem, quoted above, also the restriction of the convex surfaces F, F' is unnecessary, since Herglotz's result [4], as predicted by Weyl [7], has nothing to do with the analyticity, but only with a specific degree of differentiability (C^n), of the given pair of surfaces; cf. Section 4 below. But then the theorem has again two interpretations, one being the above (b) and another, say (a_n), an interpretation which would be an appropriate C^n-analogue of the above (a); cf. the end of Section 4 below.

3. There is however a classical instance of isometry in which, in contrast to the Weyl-Herglotz problem, the analyticity of the surfaces is precisely the issue. It is S. Bernstein's theorem, the statement of which runs as follows: "If S and an analytic S' of positive Gaussian curvature are isometric, then S is analytic." This is Bernstein's own formulation of his theorem (the italicized statement in [1], p. 434). But it is again not specified which of the two interpretations defined above, (a) or (b), is meant under "isometry," and so it is again necessary to point out that the weaker formulation, (b), implies the stronger one, (a); see Section 3 below.

In addition, the above wording of Bernstein's theorem fails to specify the degree of smoothness required of the given surface, S. In this regard, Bernstein's proof makes it clear that S is assumed to be of class C^3, rather than, as one might desire (and expect from the above wording), of class C^1 only. This comes about by Bernstein's use of his general theorem. according to which every function $z = z(x, y)$ satisfying a partial differential equation, of second order, of elliptic type, and having analytic coefficients, must be analytic whenever it is of class C^3. Accordingly, the improved version of the above wording of Bernstein's theorem is as follows:

If S is of class C^3 and of positive Gaussian curvature, and if there exists a C^1-transformation (4) *of S into an analytic S': $X'(u'^1, u'^2)$, satisfying* (5) *(where the parametrization S: $X(u^1, u^2)$ is of class C^3), then S is analytic.*

Bernstein's proof is based on an application of geodesic polar coordinates

([1], p. 435). But a perusal of his proof shows that the application of such particular (u^1, u^2)-parameters is unimportant. In fact, under the assumption that S has a C^1-parametrization (1), all that is needed is the existence of a C^1-transformation (4) satisfying (5) and having the property that, while the vector function $X(u^1, u^2)$ is supposed to be of class C^1 in (u^1, u^2), it becomes of class C^3 in (u'^1, u'^2) by virtue of (4). The rest then follows from Bernstein's general theorem on elliptic differential equations and from Section 4 below.

These remarks also show that the theorem can well be formulated so as to avoid the difficulties involved in the various concepts of an isometry; namely, as follows:

Let (g_{ik}) be a binary, symmetric matrix of analytic functions

$$g_{ik} = g_{ik}(u^1, u^2)$$

and suppose that the curvature [5] of

$$ds^2 = g_{ik}(u^1, u^2)\, du^i du^k \tag{6}$$

is positive. Then a vector function $(x, y, z) = X = X(u^1, u^2)$ satisfying

$$| dX(u^1, u^2)|^2 \equiv ds^2 \tag{7}$$

must be analytic whenever it is of class C^3.

While this theorem might be true even if the last C^3 is relaxed to C^1, its truth or falsehood is undecided even if the C^3 is relaxed just to C^2 (in which latter case, but not in the C^1-case, (1) must have a curvature not only by virtue of the Theorema Egregium but in terms of the normal image of S as well).

4. If S and S' are of class C^n [analytic], let all their parametrizations (3) considered be restricted to C^n-parametrizations [analytic parametrizations], and let (4) be called a C^m-isometry [an analytic isometry] of the pair (3) if (5) is satisfied and (4) is a C^m-transformation [an analytic transformation] (in the sense that both functions occurring in (4) are of class C^m [analytic] with a non-vanishing Jacobian).

It is understood that, in the non-analytic case of this definition, $1 \leqq n \leqq \infty$ and $1 \leqq m \leqq \infty$, and that $m \neq n$ is allowed. In the analytic case, the above critique (Sections 2-3) will be disposed of by the following lemma:

If S and S' are analytic and C^1-isometric, then all of their C^1-isometries are analytic isometries.

---

[5] Calculated from the Theorema Egregium.

This lemma is a corollary of (i) the fact that, according to Gauss [2], every analytic S admits of an analytic parametrization (1) in which the line-element (6) appears in the normal form

$$ds^2 = g(u^1, u^2)\{(du^1)^2 + (du^2)^2\}, \qquad (g = g_{ii} > 0, g_{12} = 0);$$

to be combined with (ii) the fact that if two real-valued functions (4) are of class C^1 and satisfy the Cauchy-Riemann equations, then $u'^1 + iu'^2$ is an analytic function of $u^1 + iu^2$.

First, since S and S' are analytic, two applications of (i) supply the existence of two analytic transformations (of non-vanishing Jacobian), say of

$$\alpha\colon\ (u^1, u^2) \to (v^1, v^2) \text{ and } \beta\colon\ (u'^1, u'^2) \to (v'^1, v'^2),$$

by virtue of which the respective line-elements $|\, dX(u, v)\,|^2$ and $|\, dX'(u', v')\,|^2$ on S and S' become of the form

$$ds^2 = h(v^1, v^2)\{(dv^1)^2 + (dv^2)^2\} \text{ and } ds'^2 = h'(v'^1, v'^2)\{(dv'^1)^2 + (dv'^2)^2\},$$

where $h > 0$, $h' > 0$. On the other hand, since S and S' are supposed to be C^1-isometric, there exist two functions (4), of class C^1 and of non-vanishing Jacobian, satisfying

$$ds^2 = ds'^2 \text{ by virtue of } \gamma\colon\ (u^1, u^2) \to (u'^1, u'^2).$$

But the last three formula lines show that $\beta^{-1}\gamma\alpha^{-1}$ is a conformal transformation of a domain in the Euclidean (v^1, v^2)-plane into a domain in the Euclidean (v'^1, v'^2)-plane, and $\beta^{-1}\gamma\alpha^{-1}$ is a C^1-transformation, since γ is. It follows therefore from (ii) (where every u must be replaced by the corresponding v) that $\beta^{-1}\gamma\alpha^{-1}$ is analytic. In view of the analyticity of β and α, this proves that γ is analytic, which is the assertion of the above lemma.

One might think that this proof also leads to a C^n-analogue of the above lemma, with "conformal" transformations α and β which, instead of being analytic, as above, are of class C^n. Actually, this approach fails, since it applies only to the classes (2) (with an unspecified λ). In fact, Lichtenstein's analogue [6] of the above (i) is as follows: If $n > 0$ and $0 < \lambda^* < \lambda < 1$, then every S of class $C^n(\lambda)$ can conformally be mapped on a domain in the Euclidean plane by a transformation (4) of class $C^n(\lambda^*)$. But it is not known (and it is probably not true; cf. Section 5 below) that this remains true if both $C^n(\lambda)$ and $C^n(\lambda^*)$ are replaced by C^n itself.[7] Nevertheless, the straight

---

[6] Cf. [5], where the proof is given for $n = 1$ only but, as easily realized, applies for $n > 1$ also.

[7] Or, for that matter, if $0 < \lambda^* < \lambda < 1$ is replaced by $0 < \lambda^* = \lambda < 1$.

C^n-analogue of the above analytic lemma might be true, since all that follows is the failure of *conformal* normal forms.

5. The question raised at the end of Section 4, concerning the necessity of a Hölder index λ, can slightly be generalized, by omitting the restriction that the binary Riemannian line-element (6) be "embedded" in the X-space, as required by (1) and (7). Then, if (u^1, u^2), (v^1, v^2) are denoted by (u, v), (p, q), and g_{11}, g_{12}, g_{22} by E, F, G, respectively, the question becomes, for every fixed positive n (including $n-1=0$) the following:

($?_n$) If E, F, G are given functions of class C^{n-1} on a sufficiently small $(u. v)$-domain, and if

$$EG - F^2 > 0 \tag{8}$$

(that is, if the quadratic form

$$ds^2 = E(u,v)\,du^2 + 2F(u,v)\,du dv + G(u,v)\,dv^2 \tag{9}$$

is definite), then must there exist two functions

$$p = p(u,v), \qquad q = q(u,v), \tag{10}$$

of class C^n and of non-vanishing Jacobian, with the property that the form (9) becomes

$$ds^2 = D(p,q)\,(dp^2 + dq^2) \tag{11}$$

by virtue of (10) (for some function $D > 0$, determined by the Jacobian of (10))?

The answer to ($?_1$) is affirmative if and only if (8) and the mere continuity ($C^{n-1} = C^0$) of the three coefficient functions of (9) are always sufficient to assure the existence of at least one C^1-solution (10) (of non-vanishing determinant) for Beltrami's form of the Cauchy-Riemann equations, that is, for

$$q_u = (Fp_u - Ep_v)/W, \qquad q_v = -(Fp_v - Gp_u)/W, \tag{12}$$

where $W^2 = EG - F^2$.

If a solution (10) of (12), instead of being of class C^1 only, is of class C^2, then, since the rule $\phi_{uv} = \phi_{vu}$ becomes applicable to both functions (10), it follows from (12), and from the non-vanishing of the Jacobian of (10), that both $\phi = p$ and $\phi = q$ are solutions $\phi = \phi(u,v) \neq \text{const.}$ of

$$\{(G\phi_u - F\phi_v)/(EG - F^2)^{\frac{1}{2}}\}_u + \{(E\phi_v - F\phi_u)/(EG - F^2)^{\frac{1}{2}}\}_v = 0 \tag{13}$$

(Laplace-Beltrami). Hence, if the answer to the case $n=2$ of the question $(?_n)$ is affirmative, then the (homogeneous, linear, elliptic) partial differential equation (13) must possess some non-constant solution $\phi=\phi(u,v)$ of class C^2 whenever $E(u,v)$, $F(u,v)$, $G(u,v)$ are functions of class C^1 satisfying (8) (incidentally, this can be concluded without a detour through (12) also; cf. [6], pp. 1295-1297). But it is unlikely that this (hence an affirmative answer to the question $(?_n)$ if $n=2$) should turn out to be true. The reason for being skeptical is as follows:

Choose

$$E(u,v)\equiv 1 \text{ and } F(u,v)\equiv 0 \tag{14}$$

(so that u, v are, in the main, "geodesic polar coordinates" in the sense of Gauss, with reference to the metric (9) and to a point of the (u,v)-plane). It is clear from (14) that (8) is satisfied if $G>0$, that (12) reduces to

$$p_v=-g(u,v)q_u, \qquad q_v=g(u,v)p_u, \tag{15}$$

if g denotes $G^{\frac{1}{2}}>0$, and that (13) therefore simplifies to

$$(\phi_u g)_u+(\phi_v/g)_v=0. \tag{16}$$

Hence, if the answer to the question $(?_2)$ is affirmative, then (16) must have a solution $\phi=\phi(u,v)\neq$ const. of class C^2 whenever $g=g(u,v)$ is a positive function of class C^1. But then the coefficients of (16), being composed of g_u, g_v and g, are just continuous, and so (16) does not seem to be substantially different from the differential equation

$$\phi_{uu}+\phi_{vv}+f(u,v)\phi=0, \tag{17}$$

in which the given function, f, is just continuous. However, it was shown in [8] that it is possible to choose a continuous function $f(u,v)>0$ in such a way that (17) will fail to have *any* (continuous) solution $\phi(u,v)\neq$ const. on *any* (u,v)-domain.

6. Since the answer to the questions $(?_n)$ is not known, it is worth mentioning that the answer is surely in the negative if (8) is replaced by

$$EG-F^2<0 \tag{18}$$

and, correspondingly, (11) by

$$ds^2=D(p,q)(dp^2-dq^2). \tag{19}$$

For the case $n=1$ of the respective class C^{n-1}, C^n (a case in which E, F, G

are just continuous, hence (10) is just a C^1-transformation), this can, in other words, be formulated as follows: There exist *hyperbolic line-elements* (9), *with continuous coefficient functions* E, F, G, *which cannot be mapped "conformally" on the non-Euclidean* (p, q)*-plane* by any mapping (10) of class C^1.

In order to prove this, choose (9) as in (14). Then (18) reduces to $G < 0$ and, since (11) is replaced by (19), what corresponds to the Cauchy-Riemann system (15) becomes

$$p_v = f(u, v) q_u, \qquad q_v = f(u, v) p_u, \tag{20}$$

where $f = (-G)^{\frac{1}{2}} > 0$. Hence the assertion is that, if $f(u, v)$ is an arbitrary positive continuous function, then (20) need not have any solution (10) of class C^1, unless the Jacobian $\partial(p, q)/\partial(u, v)$ vanishes identically.

In order to prove the existence of such an f, write (20) in the form $(p \pm q)_v = \pm f(u, v)(p \pm q)_u$. The latter shows that $\phi = p + q$ must satisfy

$$\phi_v - f(u, v) \phi_u = 0 \tag{21}$$

(and $\phi = p - q$ the homogeneous, linear, partial differential equation which results from (21) if f is replaced by $-f$). But it was shown in [8], pp. 733-734. that there exist continuous functions $f = f(u, v) > 0$ having the property that, no matter where, and no matter how small, a (u, v)-circle be chosen, the differential (21) will not possess within the circle any solution $\phi = \phi(u, v)$ of class C^1, except the trivial solution $\phi(u, v) = \text{const.}$ In view of the connection between (21) and (20), this proves more than what was claimed above for the case $n - 1 = 0$.

7. The existence of an f, satisfying the conditions used in connection with (20), has a significance from the point of view of the theory of multipliers (Euler). In order to see this, let P and Q be continuous on a simply connected open (x, y)-domain, and let J denote any smooth Jordan curve, finally D any open set contained in this domain. If $\mu(x, y)$ is a continuous function on the latter, and if the line integral, along every C, of $\mu(x, y)$ times

$$P(x, y)\,dx + Q(x, y)\,dy \tag{22}$$

vanishes, then $\mu(x, y)$ will be called a multiplier of the Pfaffian (22), provided that $\mu(x, y)$ does not vanish identically. It turns out that, if no assumptions of smoothness (involving partial derivatives) are placed on the functions $P(x, y)$, $Q(x, y)$, then *it is possible to choose the Pfaffian* (22), *with con-*

tinuous $P(x, y)$ *and* $Q(x, y)$, *in such a way that on no open* (x, y)*-domain* D *will* (22) *possess a multiplier.*

In order to see this, suppose that (22) is such as to possess a multiplier, $\mu(x, y)$. Then the line integral of $\mu(x, y)$ times the Pfaffian (22), when extended from a fixed (x_0, y_0) to a variable (x, y) along a smooth Jordan arc, will be a function of (x, y) and, if this function is denoted by $\phi(x, y)$, the partial derivatives ϕ_x, ϕ_y will exist and satisfy the relations

$$(23) \qquad \phi_x(x, y) = \mu(x, y)P(x, y), \qquad \phi_y(x, y) = \mu(x, y)Q(x, y).$$

In addition, ϕ_x and ϕ_y are continuous. This follows from (23), since P, Q and μ are supposed to be continuous. Accordingly, $\phi(x, y)$ is of class C^1.

Choose $P(x, y) = 1$ for every (x, y), put $Q = f$, and write u and v instead of y and x, respectively. Then (23) becomes

$$(24) \qquad \phi_v(u, v) = \mu(u, v), \qquad \phi_u(u, v) = \mu(u, v)f(u, v)$$

(if u and v are interchanged as arguments). Substitution of $\mu(u, v)$ from the first of the relations (24) into the second shows that $\phi(u, v)$ is a solution of (21). Hence, if $f(u, v)$ is so chosen as at the end of Section 6, then, since $\phi(u, v)$ is of class C^1, it follows that $\phi(u, v) = \text{const.}$ on every D.

In particular, the partial derivative $\phi_v(u, v)$ vanishes identically. In view of the first of the relations (24), this means that μ vanishes identically. Since such a μ was excluded in the definition of a multiplier, the proof is complete.

8. In Section 2, reference was made to Herglotz's theorem [4], which states that, under certain assumptions of smoothness, two closed, essentially convex [8] surfaces are congruent whenever they are isometric ("congruence" is meant in the sense including "anti-congruence," i. e., reflections on a plane of the Euclidean X-space are allowed). With regard to the assumptions of smoothness, the situation is as follows: While the so-called derivation formulae (those of Gauss and Weingarten) hold on any surface of class C^2, Herglotz's proof involves differentiations (of first order) of these formulae and assumes therefore that the convex surfaces are of class C^3. In addition, the proof depends on a tacit assumption, one corresponding to the comments in Sections 1-2 above. In fact, the differentiated formulae, just mentioned, contain local

[8] A convex surface (of class C^2) will be called essentially convex if the set of its parabolic points (if there are any) is of Lebesgue measure 0; cf. (52) below. For instance, convexity implies essential convexity under the proviso that the Gaussian curvature should be positive with the possible exception of isolated points or curves.

representations of the two C^3-surfaces in terms of the *same* parameter plane (u^1, u^2). According to the terminology introduced at the beginning of Section 4, this additional hypothesis means that the two surfaces are supposed to be locally C^3-isometric.

From a geometrical point of view, there is an objection to most C^3-assertions in the theory of surfaces. In fact, the curvatures (total and mean, K and H) and even both fundamental forms ($g_{ik}du^i du^k$ and $h_{ik}du^i du^k$) exist, and are continuous, on surfaces of class C^2. Correspondingly, the restriction of a theorem to surfaces of class C^3, instead of to more inclusive (and *geometrical*) class C^2, is often due to accidental formal difficulties, resulting from the limitations of the underlying analytical tools, rather than to the actual geometrical situation.

For instance, while the fundamental existence theorem of the differential geometry of surfaces (Bonnet) is a C^3-theorem in its classical wording, it can, with some effort, be freed (cf. [3], pp. 758-760) of the unnatural C^3-restriction. In what follows, the possibility of a corresponding reduction, $C^3 \rightarrow C^2$, will be proved in Herglotz's theorem, both with regard to the smoothness of the two convex surfaces and that of their underlying isometry. In other words, it will be proved that the content of the theorem can be refined to the following statement:

Two closed, essentially[9] *convex surfaces of class C^2 are congruent whenever they are (locally) C^2-isometric.*

It would be desirable to reduce the theorem even further, by showing that every (local) C^1-isometry of two C^2-surfaces is a C^2-isometry by necessity (cf. Section 4).

9. As in (3), let S and S' be two, sufficiently small, pieces of surfaces both of which are of class C^2. While in the application of the lemma to be derived, S and S' represent pieces of two closed, convex surfaces, no such additional assumption is made now (so that the Gaussian curvatures need not be non-negative). Suppose that S and S' are C^2-isometric. Then after a suitable C^2-transformation (4) (of non-vanishing Jacobian), (3) can be assumed to be in the form

$$S\colon X(u^1, u^2) \text{ and } S'\colon X'(u^1, u^2), \tag{25}$$

where

$$| dX(u^1, u^2) |^2 = | dX'(u^1, u^2) |^2 \tag{26}$$

---

[9] See the preceding footnote.

(since (4) in (5) is now $u'^1 = u^1$, $u'^2 = u^2$), and where the vector functions (25) are of class C^2 and such that the vector products $[X_1, X_2]$, $[X'_1, X'_2]$ of the respective partial derivatives $Z_i = \partial Z/\partial u^i$ ($i = 1, 2; Z = X, X'$) do not vanish. Thus there exist normal unit vectors, say

(27) $\quad N = [X_1, X_2]/|[X_1, X_2]|$ and $N' = [X'_1, X'_2]/|[X'_1, X'_2]|$,

and, according to (26), the scalar products

(28) $$g_{ik} = X_i \cdot X_k, \qquad g_{ik} = X'_i \cdot X'_k$$

are respectively identical (and have a determinant which is positive, say

(29) $$g > 0, \text{ where } g = (\det g_{ik})^{\frac{1}{2}},$$

since $[X_1, X_2] \neq 0$), and the functions (27), (28), (29) of (u^1, u^2) are of class C^1. In contrast to (28), expressing the identity of the two first fundamental forms, $g_{\alpha\beta}du^\alpha du^\beta$ and $g'_{\alpha\beta}du^\alpha du^\beta$, there are two distinct second fundamental forms, the coefficients of which are defined by

(30) $$h_{ik} = N \cdot X_{ik}, \qquad h'_{ik} = N' \cdot X'_{ik}$$

($X_{ik} = \partial^2 X/\partial u^i \partial u^k$), and, since the functions (25) and (27) are of class C^2 and C^1, respectively, the functions (30) are just continuous.

The two matrices (h_{ik}), (h'_{ik}) (which can be definite, semi-definite or indefinite) have the same determinant at every point (u^1, u^2); in other words,

(31) $$g^2 K = \det h_{ik}, \qquad g^2 K = \det h'_{ik},$$

if (29) and the first of the relations (31) are considered as the definition of the continuous function $K = K(u^1, u^2) \gtreqless 0$. If the function (25) were of class C^3, hence the functions (28) of class C^2, then, since (31) defines the Gaussian curvatures, K, of S and S', respectively, the identity claimed by (31) would follow from the classical form of the Theorema Egregium (note that the latter contains the second derivatives of functions g_{ik}). This proof of (31) fails to apply, since the functions g_{ik} are supposed to be of class C^1 only. That (31) is nevertheless true in the present case also, follows from the circumstance that, due to a fact first observed by Weyl, a certain integrated form of the Theorema Egregium happens to hold for every surface of class C^2 (for references and for a simple proof, cf. [3], p. 760 and formula (7) on p. 759).

Besides the (common) Gaussian curvature, K, of S and S', consider their (generally different) mean curvatures,

(32) $\quad H = \frac{1}{2} g^{\alpha\beta} h_{\alpha\beta}. \qquad$ (33) $\quad H' = \frac{1}{2} g^{\alpha\beta} h'_{\alpha\beta}.$

where (g^{ik}) denotes the reciprocal matrix, $(g_{ik})^{-1}$, of (g_{ik}). Put

$$2g^2J = h_{11}h'_{22} - 2h_{12}h'_{12} + h_{22}h'_{11}. \tag{34}$$

In view of (29), the function $J = J(u^1, u^2)$ defined by (34) is the "mixed" form (in the sense of Brunn and Minkowski) of the two expressions (31). (This "mixed" Gaussian curvature seems first to have arisen in connection with Weingarten's "associated surfaces" of his theory of infinitesimal deformations; cf. the reference in footnote [10] below.) All three functions (32), (33), (34) are continuous, since the functions (30), (28) are.

For the above-defined functions, the integral relation

$$\int_C g^{-1}(h'_{2\alpha}X \cdot X_1 - h'_{1\alpha}X \cdot X_2)du^\alpha = 2\iint_D g\{(N \cdot X)J + H'\}du^1du^2 \tag{35}$$

is an identity in C, where C on the left denotes a piecewise smooth, Jordan curve contained in the (sufficiently small and hence, without loss of generality, simply connected) parametric $(u^1. u^2)$-domain on which the surfaces (25) are given, while D on the right of (35) denotes the interior of C. This is Herglotz's fundamental identity, proved by him under his C^3-assumption for (25). It will be shown that (35) holds under the present C^2-assumption also. The C^2-theorem on closed, convex surfaces, as announced in Section 8, will then follow from the "local" identity (35), since the balance of Herglotz's proof remains unaltered.

10. The derivation formulae of Gauss and Weingarten,

$$X_{ik} = \Gamma^\alpha_{ik}X_\alpha + h_{ik}N \tag{36}$$

and

$$N_i = -g^{\alpha\beta}h_{i\alpha}X_\beta \tag{37}$$

(cf. [3], p. 758), where the $\Gamma^i_{jk} = \Gamma^i_{jk}(u^1, u^2)$ are Christoffel's symbols, hold on every surface of class C^2. On the other hand, the Mainardi-Codazzi equations (representing that part of the integrability conditions of the system (36)-(37) which remains after a satisfaction of the Theorema Egregium) cannot be applied on a surface of class C^2, since they contain the derivatives of the functions h_{ik}, whereas the latter functions are of class C^1 only if the surface is of class C^3. It was however shown in [3], pp. 759-760, that if the surface is of class C^2, then the Mainardi-Codazzi equations still apply in their "integrated" form,

$$\int_C h_{i\alpha}du^\alpha = \iint_D (\Gamma^\alpha_{i1}h_{\alpha 2} - \Gamma^\alpha_{i2}h_{\alpha 1})du^1du^2, \qquad (i = 1, 2), \tag{38}$$

14

where the ("arbitrary") Jordan curve C and its interior D are restricted only by the assumptions which were specified for (35) above.

Let (36′), (37′), (38′) denote the relations which result from (36), (37), (38) if X, N, h_{ik} are replaced by X', N', h'_{ik}, respectively (note that $\Gamma' = \Gamma$, since $g'_{ik} = g_{ik}$).

Since the function (29) is of class C^1, an application of the general Lemma of [3], p. 761, shows that (38) implies the relations

$$\int\limits_C g^{-1}h'_{ia}du^a = \int\!\!\int\limits_D g^{-1}\Delta_i du^1 du^2, \qquad (i = 1, 2), \tag{39}$$

if the (continuous) functions Δ_1, Δ_2 are defined by

$$\Delta_i = (-1)^j(\Gamma^j{}_{11}h'_{22} - 2\Gamma^j{}_{12}h'_{12} + \Gamma^j{}_{22}h'_{11}), \qquad j \neq i \tag{40}$$

(so that $j = 2$ or $j = 1$ according as $i = 1$ or $i = 2$). The same general lemma also shows that the partial derivative, $x_2 = \partial x/\partial u^2$, of the first component of the vector $X = (x, y, z)$ can be "inserted" into the case $i = 1$ of the relation (39), and that this leads to

$$\int\limits_C g^{-1}x_2h'_{1a}du^a = \int\!\!\int\limits_D g^{-1}(\Delta_1x_2 + h'_{12}x_{12} - h'_{11}x_{22})\,du^1du^2.$$

Similarly,

$$\int\limits_C g^{-1}x_1h'_{2a}du^a = \int\!\!\int\limits_D g^{-1}(\Delta_2x_1 + h'_{22}x_{11} - h'_{12}x_{12})\,du^1du^2.$$

Hence, by subtraction,

$$\int\limits_C g^{-1}(x_1h'_{2a}du^a - x_2h'_{1a}du^a) = \int\!\!\int\limits_D [\cdot\ \cdot\ \cdot]\,du^1du^2, \tag{41}$$

where the expression on the right is the difference of the double integrals occurring in the preceding two formulae.

Corresponding to $X = (x, y, z)$, let $N = (a, b, c)$. Then, according to (36),

$$x_{ik} = \Gamma^a{}_{ik}x_a + ah_{ik}.$$

Hence, if x_{11}, x_{12}, x_{22} are multiplied by $g^{-1}h'_{22}$, $-2g^{-1}h'_{12}$, $g^{-1}h'_{11}$, respectively, it is seen, by addition, that the difference which on the right of (41) was indicated by $[\cdot\ \cdot\ \cdot]$ can be written as

$$[\cdot\ \cdot\ \cdot] = g^{-1}a(h_{11}h'_{22} - 2h_{12}h'_{12} + h_{22}h'_{11}),$$

since the functions Δ_1, Δ_2 occurring in the definition of the difference $[\cdot\cdot\cdot]$ are given by (40).

According to (34), the preceding representation of $[\cdot\cdot\cdot]$ can be written as $[\cdot\cdot\cdot]=2gaJ$. Thus (41) becomes

$$(42)\qquad \int_C g^{-1}(x_1h'_{2\alpha}du^\alpha - x_2h'_{1\alpha}du^\alpha) = \iint_D 2gaJ\,du^1du^2,$$

where a denotes the direction cosine defined by $N=(a,b,c)$. If the general Lemma of [3], p. 761, is applied again, this time in order to "insert" x as a factor gaJ on the right of (42), it follows that (42) implies the relation

$$\int_C g^{-1}x(x_1h'_{2\alpha}du^\alpha - x_2h'_{1\alpha}du^\alpha)$$

$$=\iint_D \{2gaxJ + g^{-1}[x_1(x_1h'_{22}-x_2h'_{12}) - x_2(x_1h'_{12}-x_2h'_{11})]\}du^1du^2.$$

But x, a can here be replaced by any of the respective components of $X=(x,y,z)$, $N=(a,b,c)$. Hence the last relation, when compared with the definitions (28) and (32), completes the proof of (35).

11. It was mentioned at the end of Section 9 that it is sufficient to prove (35) under the C^2-assumption in (25). Actually, since the end of Herglotz's argument depends on an appeal to the fundamental uniqueness theorem of local differential geometry, and since the classical formulation of this theorem (Bonnet) is confined to surfaces of class C^3, for the sake of completeness the balance of the proof will also be given. (Incidentally, it will be worth noting that, just as the Gauss-Bonnet representation of the genus, the classical formula for the surface average of the mean curvature and Herglotz's generalization of this formula for the "mixed" case, namely (48) and (44) below, are not restricted to surfaces of genus 0; cf. [4], p. 128.)

Let F and F' be two orientable, closed, homeomorphic, locally C^2-isometric surfaces of class C^2. Draw on F an oriented net, and on F' the corresponding net, of piecewise smooth Jordan curves in such a way that, if $S_1,\cdots,S_h$ and $S'_1,\cdots,S'_h$ denote the interiors of these (oriented) Jordan curves, then, on the one hand, every pair $(S,S')=(S_j,S'_j)$, where $j=1,\cdots,h$, has a C^2-representation of the form (25) and, on the other hand, (35) holds for every $C=C_j$ and for the corresponding $D=D_j$, where D_j denotes the (u^1,u^2)-domain to which S_j and S'_j are referred in (25). Then, if (35) is applied to every j, summation with respect to j gives

$$0 = \iint_E g\{(N \cdot X)J + H'\} du^1 du^2,$$

where $E = D_1 + \cdot \cdot + D_h$. This step assumes that the line integrals, which are cancelled by the addition, have a geometrical meaning (in the sense that they are independent of the different parametrizations, used on the different pieces S_j), which, however, can easily be ascertained.

Let dF and dF' denote the surface elements on F and F', respectively, and put

$$p = N \cdot X. \tag{43}$$

Then the preceding integral relation can be written in the form

$$\iint_F J p dF = -\iint_{F'} H' dF', \tag{44}$$

since, according to (29) and (28), both dF and dF' are identical with $g du^1 du^2$. It follows that

$$\iint_F L p dF = \iint_{F'} H' dF' - \iint_F H dF, \tag{45}$$

if L is defined by

$$2g^2 L = \det(h'_{ik} - h_{ik}) \qquad (g^2 = \det g_{ik} > 0). \tag{46}$$

In fact, it is seen from (31)-(33) and (46) that

$$K - J = L. \tag{47}$$

But if F' is particularized to F, then (46) reduces to $L = 0$, hence (47) to $K = J$, and therefore (44) to

$$\iint_F K p dF = -\iint_F H dF. \tag{48}$$

Finally, (45) follows by subtracting (44) from (48) and using (47).

12. The closed, orientable surfaces F, F' have thus far been of arbitrary genus. It will be supposed that the Gaussian curvature $K = K(u^1, u^2)$ is positive throughout or, more generally, that

$$K > 0 \text{ almost everywhere} \tag{49}$$

(hence $K \geqq 0$ everywhere) on F and/or F'. Then, according to Hadamard, the genus must be 0, and so F, F' are closed convex surfaces. In particular, since (43) is the function of support for F, it can be assumed, (by choosing the origin of the X-space at a point which is not on F) that $p = p(u^1, u^2)$ is positive at every point of F.

A repetition of the argument used by Herglotz now shows that

$$(50) \qquad h'_{ik} = h_{ik}$$

holds as an identity. His proof of this identity can be modified as follows: [10] It is readily verified that if $a_{\alpha\beta}x^\alpha x^\beta$, $b_{\alpha\beta}x^\alpha x^\beta$ are two positive definite, binary, quadratic forms of common determinant, $\det a_{ik} = \det b_{ik} > 0$, then their difference is either an indefinite or the null form (i. e., either $\det(a_{ik} - b_{ik}) < 0$ or $(a_{ik}) = (b_{ik})$ must hold).

In view of (49) and (31), the assumptions of this alternative are satisfied by $a_{ik} = h_{ik}(u^1, u^2)$, $b_{ik} = h'_{ik}(u^1, u^2)$ at almost every point (u^1, u^2). It follows therefore from (49) that, if a (u^1, u^2)-set of measure 0 is disregarded, then either $L < 0$ or (50) holds at each of the remaining points (in particular, $L \leqq 0$ holds everywhere). Hence it is seen from (49) and (45), where $p > 0$, that

$$(51) \qquad 0 \geqq \iint_{F'} H'dF' - \iint_{F} HdF.$$

For reasons of symmetry, (51) must remain true if F and F' are interchanged. Consequently, the last inequality must actually be an equality. It follows therefore from (45) that

$$(52) \qquad \iint_{F} LpdF = 0.$$

Since $L \leqq 0$ and $p > 0$ hold everywhere, (52) implies that $L = 0$ holds almost everywhere, and therefore, by continuity, everywhere. In view of (46), this proves (50).

In order to complete the proof of the theorem, it is only necessary to apply to (28) and (50) the local uniqueness theorem of [3] (Theorem (I), p. 760), which states that the first and second fundamental forms of a surface S *of class* C^2 determine S uniquely.

THE JOHNS HOPKINS UNIVERSITY.

---

[10] This variant of the corresponding explicit calculation in Herglotz's paper seems to be well-known. It was used by Professor Heinz Hopf in a lecture given this spring at Johns Hopkins, and it was known to the writer since he first read Herglotz's paper. Cf. also L. Bianchi, *Vorlesungen über Differentialgeometrie* (ed. 1899), the last quarter of p. 293 and formula (8*).

REFERENCES.

[1] S. Bernstein, "Sur la déformation des surfaces," *Mathematische Annalen*, vol. 60 (1905), pp. 434-436.

[2] C. F. Gauss, *Werke*, vol. 4 (1873), pp. 193-216.

[3] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[4] G. Herglotz, "Ueber die Starrheit der Eiflächen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 15 (1943), pp. 127-129.

[5] L. Lichtenstein, "Zur Theorie der konformen Abbildung. Konforme Abbildung nichtanalytischer, singularitätenfreier Flächenstücke auf ebene Gebiete," *Bulletin International de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques et Naturelles*, ser. A, 1916, pp. 192-217.

[6] ———, "Neuere Entwicklung der Theorie partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Encyklopädie der Mathematischen Wissenschaften*, II C 12 (1924), pp. 1277-1334.

[7] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[8] A. Wintner, "On the Hölder restrictions in the theory of partial differential equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 731-738.

ON UNSMOOTH TWO-DIMENSIONAL RIEMANNIAN METRICS.*

By Philip Hartman.**

This paper will be concerned with two quite different questions. Part I will deal with the question of the embedding into 3-dimensional Euclidean space a 2-dimensional Riemannian metric which has the Tchebychef normal form $du^2 + 2\cos\phi\, dudv + dv^2$, where it is assumed that this form has the "curvature" -1, but it is only assumed that ϕ is continuous. It will be shown that such an embedding exists and is unique, up to Euclidean movements, when $u = \text{const.}$ and $v = \text{Const.}$ are required to be asymptotic lines.

In Part II, it will be shown that if a positive-definite form $g_{ik}du^i du^k$, where $i, k = 1, 2$, of class C' is transformed into another form $G_{ik}dU^i dU^k$ with the same properties by a transformation $u^i = u^i(U^1, U^2)$ of class C', then the transformation is necessarily of class C''. This result is used to establish the uniqueness statement of Part I and has other applications, cf. § 7 below.

Part I.

1. Let $z = z(x, y)$, defined in a vicinity of $(x, y) = (0, 0)$, be a surface S of class C'' and possess the Gaussian curvature $K = -1$. It has been shown [4] that these conditions on S are sufficient to assure the existence of a transformation, defined in a vicinity of $(u, v) = (0, 0)$,

$$(1) \qquad x = x(u, v), \qquad y = y(u, v), \qquad (x(0,0) = y(0,0) = 0),$$

of class C', with non-vanishing Jacobian and with the properties that, in the resulting parametrization $X = X(u, v) = (x(u, v), y(u, v), z(x(u, v), y(u, v)))$ of S, the arcs $u = \text{const.}$ and $v = \text{Const.}$ are asymptotic curves and the squared element of arc-length on S has the Tchebychef form

$$(2) \qquad ds^2 = g_{ik}(u, v)du^i du^k = du^2 + 2\cos\phi\, dudv + dv^2, \qquad (u^1, u^2) = (u, v).$$

Since $\det(g_{ik}) = \sin^2\phi$, it follows that

$$(3) \qquad \sin\phi \neq 0, \qquad (\phi = \phi(u, v)).$$

* Received February 1, 1951.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

Finally, the function ϕ satisfies the Hazzidakis relation

(4) $$[\phi] = \int_{u_1}^{u_2}\int_{v_1}^{v_2} \phi(u,v)\,dudv, \text{ where } u_1 \leqq u \leqq u_2, v_1 \leqq v \leqq v_2$$

is any rectangle on which (1) is defined and

(5) $$[\phi] = \phi(u_1, v_1) - \phi(u_2, v_1) + \phi(u_2, v_2) - \phi(u_1, v_2).$$

Since (1) is of class C', the function ϕ is continuous. It was also shown in [4] that ϕ is of class C' if and only if $z = z(x, y)$ is of class C'''. (The existence of surfaces $S: z = z(x, y)$ which are of class C'', without being of class C''', and which are pseudo-spheres ($K = -1$) will be clear from the considerations below.)

In this paper, a converse of the above result will be considered.

(*) *Let* $\phi = \phi(u, v)$ *be a continuous function on the rectangle*

(6) $$R: |\, u \,| \leqq a, \quad |\, v \,| \leqq b,$$

satisfying the inequality $0 < \phi < \pi$ *and the relation* (4) *for all rectangles in* (6). *Then, for sufficiently small* x, y, *there exists one and, up to Euclidean movements of the* (x, y, z)*-space, only one pseudo-sphere* $S: z = z(x, y)$ *of class* C'' *which belongs to* $\phi(u, v)$ *in the sense of the paragraph above.*

2. In order to make clear the content of this assertion, suppose first that ϕ is of class C', then (4) is equivalent to the existence and continuity of the second mixed partial derivative $\phi_{uv} = \phi_{vu}$ and to

(7) $$\phi_{uv} = \sin\phi.$$

Suppose further that ϕ is even smoother, say of class C'', then one expects the corresponding $X = X(u, v)$ to be of class C'''. Since $u = \text{const.}$ and $v = \text{Const.}$ are to become asymptotic arcs, the diagonal elements $h_{11}(u, v)$, $h_{22}(u, v)$ of the second fundamental matrix will be zero. Also, $K \equiv -1$ means $\det(h_{ik}) \equiv -\det(g_{ik}) = -\sin^2\phi$; so that $h_{12}{}^2 = \sin^2\phi$. Define $h_{ik} = h_{ik}(u, v)$ by

(8) $$h_{ik}du^i du^k = 2\sin\phi\, dudv.$$

The other choice, $h_{ik}du^i du^k = -2\sin\phi\, dudv$, merely corresponds to the enumeration $(u^1, u^2) = (v, u)$ (rather than to $(u^1, u^2) = (u, v)$).

It is easily verified that (g_{ik}) and (h_{ik}) given by (2) and (8), respectively, which are of class C'', satisfy the integrability conditions of Gauss and

Mainardi-Codazzi by virtue of (7). Thus, the standard theorem of Bonnet supplies a surface $S: X = X(u, v)$ of class C''' on which (2) and (8) hold. That S, in a Cartesian parametrization, say $z = z(x, y)$, is of class C'''' follows from the considerations in [4], § 14.

The variant of the theorem of Bonnet, proved in [3], shows that if ϕ is of class C', then there exists an $S: X = X(u, v)$ of class C'' on which (2) and (8) hold and which, by [4], § 14, is of class C''' in a suitable parametrization.

But if ϕ is only continuous, one cannot write down all of the equations in the linear total system, consisting of the derivation formulae of Gauss and Weingarten, for which the theorem of Bonnet and its variant supply a solution. In fact, these equations in the smooth cases of (2) and (8) are

$$(9)\quad X_{uu} = (X_u \cos\phi - X_v)\phi_u/\sin\phi, \qquad X_{vv} = (X_v \cos\phi - X_u)\phi_v/\sin\phi,$$

$$(10)\qquad X_{uv} = N \sin\phi$$

and, if $N = N(u, v)$ denotes the unit normal vector, $(X_u, X_v)/|(X_u, X_v)|$,

$$(11)\quad N_u = (X_u \cos\phi - X_v)/\sin\phi, \qquad N_v = (X_v \cos\phi - X_u)/\sin\phi,$$

while (9) involves partial derivatives of ϕ.

The existence assertion in (*) above will be proved by an approximation process. The uniqueness assertion will depend on the result of Part II, below.

3. Before beginning the proof, it can be remarked that the assertion (*) goes beyond the existence statements just mentioned; that is, there actually exist functions $\phi = \phi(u, v)$ satisfying (4), which are continuous but not of class C'. This can be proved by the existence proof of Picard for the hyperbolic differential equation (7). Let $\alpha = \alpha(u)$ and $\beta = \beta(v)$ be continuous functions for $|u| \leqq a$, $|v| \leqq b$, respectively, satisfying $\alpha(0) = \beta(0)$. Then there exists one and only one continuous function ϕ on the rectangle (6) satisfying (4) and $\phi(u, 0) = \alpha(u)$ and $\phi(0, v) = \beta(v)$. This follows by considering the successive approximations $\psi_0(u, v) = \alpha(u) + \beta(v) - \alpha(0)$ and

$$\psi_n(u, v) = \psi_0(u, v) + \int_0^u \int_0^v \sin\psi_{n-1}(u, v)\,du\,dv \quad \text{for } n = 1, 2, \cdots.$$

The standard calculation shows that $\phi = \lim \psi_n$, as $n \to \infty$, exists and is the desired function. Clearly, $\phi(u, v)$ is of class C' if and only if the given functions $\alpha(u)$, $\beta(v)$ are.

This proof also implies that if a ϕ is given, then there exists a sequence of functions $\phi^1(u,v), \phi^2(u,v), \cdots$ on (6), which are smooth, satisfy (4) (that is, (7)) and which, as $n \to \infty$, tend uniformly to $\phi(u,v)$. In fact, it is sufficient to choose sequences $\alpha^n(u)$, $\beta^n(v)$ of polynomials, which satisfy $\alpha^n(0) = \beta^n(0)$ and which approximate uniformly the respective functions $\alpha(u) = \phi(u,0)$, $\beta(v) = \phi(0,v)$, and to let $\phi^n(u,v)$ be the unique solution of (7) belonging to the initial conditions $\phi^n(u,0) = \alpha^n(u)$, $\phi^n(0,v) = \beta^n(v)$.

4. In order to prove the existence statement in (*), let $\phi = \phi(u,v)$ be a given continuous function on (16) satisfying (4) and $0 < \phi < \pi$. Let $\phi^1, \phi^2, \cdots$ be a sequence of approximating smooth functions, described above. By virtue of uniform convergence, $0 < \phi^n(u,v) < \pi$ holds for all (u,v) on (6) and for all sufficiently large n. By discarding a finite number of the $\phi^1, \phi^2, \cdots$ (and renumbering the sequence), it can be supposed that this inequality holds for all n.

Let (2_n), (7_n), (8_n), $\cdots$ denote the relations (2), (7), (8), $\cdots$, respectively, in which ϕ is replaced by ϕ^n. By the theorem of Bonnet, there exists a unique smooth surface $S^n: X = X^n(u,v)$ for which (2_n), (8_n) hold and for which

$$(12_n)\qquad \begin{aligned} &X^n(0,0) = 0, \qquad X_u{}^n(0,0) = (1,0,0), \\ &X_v{}^n(0,0) = (\cos\phi^n(0,0), \sin\phi^n(0,0), 0). \end{aligned}$$

Note that $X^n(u,v)$ exists on the entire rectangle (6), since the total system involved, (9_n)-(11_n), is linear.

Let it be granted for the moment that *the sequences* $X_u{}^1, X_u{}^2, \cdots$ and $X_v{}^1, X_v{}^2, \cdots$ *are equicontinuous on* (6). These sequences are obviously bounded since (2_n) implies $|X_u{}^n| \equiv |X_v{}^n| \equiv 1$. Hence, by the theorem of Arzelà, it is possible to select a subsequence of the surfaces $X^1, X^2, \cdots$, which will again be denoted by $X^1, X^2, \cdots$, such that $X^1, X^2, \cdots$ and its sequence of first order partial derivatives converge uniformly on (6), as $n \to \infty$. Let $S: X = X(u,v)$ denote the limit surface.

Clearly, on the surface S, the squared element of arc-length is given by (2). Furthermore, if $N = N(u,v)$ is the unit normal vector, then N is of class C'. In fact, (11) holds as a consequence of (11_n) and the selection process above. In the same way, it is seen that $X_{uv} = X_{vu}$ exists, is continuous and satisfies (10).

The initial conditions (12) imply that the Jacobian of (1) does not vanish at $(u,v) = (0,0)$, so that S has a representation of the form $z = z(x,y)$, where $z = z(x,y)$ is of class C' in a vicinity of $(x,y) = (0,0)$.

But since $N = \pm(z_x, z_y, -1)/(1 + z_x^2 + z_y^2)^{\frac{1}{2}}$ is of class C' as a function of (u, v), and hence as a function of (x, y), it follows that $z = z(x, y)$ is of class C''. The relations (4) and (10) and the theorem of Gauss-Bonnet, as applied in [4], § 15, show that the curvature of S is identically -1. Finally, a point P of S and an asymptotic direction at P determine a unique asymptotic arc on S; [2], § 1. Hence, the asymptotic arcs on S^n, $u =$ const. and $v =$ Const., tend to those of S. Consequently, the parameter lines $u =$ const. and $v =$ Const. on S are the asymptotic arcs.

5. Thus, in order to complete the existence statement contained in (*), it remains to verify the hypothesis made in the last section, that the sequences $X_u^1, X_u^2, \cdots$ and $X_v^1, X_v^2, \cdots$ are equicontinuous, by virtue of the fact that $\phi^1, \phi^2, \cdots$ is. To this end, the differences $| X_u^n(u + h, v) - X_u^n(u, v) |$, $| X_u^n(u, v + h) - X_u^n(u, v) |$ and those belonging to X_v^n will be estimated in terms of those belonging to ϕ^n (and in terms of the lower bound for $\sin \phi^n$). Since n is fixed, the notation will be simplified by omitting n.

It will be sufficient to consider the differences belonging to X_u, as those belonging to X_v can be treated similarly. Let m satisfy

(13) $$\sin \phi(u, v) \geqq m > 0 \text{ for all } (u, v) \text{ on } R.$$

If (u, v), $(u, v + h)$ are two points of R, it follows from (10) that

(14) $$| X_u(u, v + h) - X_u(u, v) | \leqq | h |$$

since $| N | = 1$.

If $(u + h, v)$, (u, v) are two points of R, let Δf denote the difference $f(u + h, v) - f(u, v)$, where f is any (scalar or vector) function on R. The difference ΔX_u will be appraised by a method similar to that used in [4], § 13.

Let C be a (fixed) number to be specified below and let $\epsilon > 0$ be any number satisfying

(15) $$2C\epsilon < m^3/4,$$

where m satisfies (13). Corresponding to ϵ, there exists a number $\delta = \delta_\epsilon > 0$ with the property that

(16) $$| \Delta\phi | < \epsilon \text{ if } | h | < \delta$$

for all (u, v), $(u + h, v)$ in R. For convenience later, it can be supposed that δ is chosen so small that

(17) $$2\delta < \epsilon m.$$

It will be shown that

$$(18) \qquad |\Delta X_u| < 8C\epsilon/m^2 \text{ if } |h| < \delta$$

for all (u,v), $(u+h,v)$ on R. If (18) is verified, then it follows from the definition (16), (17) of δ, that the existence statement in (*) is proved.

Let

$$(19) \quad A_1 = X_u(u,v) + \tfrac{1}{2}\Delta X_u, \qquad A_2 = X_v(u,v), \qquad A_3 = N(u,v).$$

If these three vectors are linearly independent, that is, if

$$(20) \qquad \det(A_1, A_2, A_3) \neq 0,$$

then there exist three unique numbers $\alpha_i = \alpha_i(u,v,h)$, where $i = 1, 2, 3$, such that

$$(21) \qquad \Delta X_u = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3.$$

If (21) is multiplied scalarly by A_1, A_2, A_3, respectively, there results a system of linear equations for α_1, α_2, α_3. It will be seen that these equations are

$$(22_1) \quad 0 = \alpha_1(A_1 \cdot A_1) + \alpha_2(A_2 \cdot A_1) + \alpha_3(A_3 \cdot A_1),$$

$$(22_2) \quad \Delta \cos \phi - X_u(u+h,v) \cdot \Delta X_v = \alpha_1(A_1 \cdot A_2) + \alpha_2(A_2 \cdot A_2) + \alpha_3(A_3 \cdot A_2),$$

$$(22_3) \quad -X_u(u+h,v) \cdot \Delta N = \alpha_1(A_1 \cdot A_3) + \alpha_2(A_2 \cdot A_3) + \alpha_3(A_3 \cdot A_3).$$

The right-hand sides of these equations are obvious. The left-hand sides can be verified as follows: the left-hand side of (22_1) is $\Delta X_u \cdot A_1 = \Delta(|X_u|^2) = \Delta(1) = 0$; the left-hand side of (22_2) is

$$\Delta X_u \cdot A_2 = \Delta(X_u \cdot A_2) - X_u(u+h,v) \cdot \Delta A_2$$
$$= \Delta \cos \phi - X_u(u+h,v) \cdot \Delta X_v;$$

finally, the left hand side of (22_3) is

$$\Delta X_u \cdot A_3 = \Delta(X_u \cdot A_3) - X_u(u+h,v) \cdot \Delta A_3$$
$$= \Delta(0) - X_u(u+h,v) \cdot \Delta N.$$

It is clear from (10), (11) and (13) that

$$(23) \qquad |\Delta X_v| \leqq |h| \text{ and } |\Delta N| \leqq 2|h|/m.$$

Hence, the terms on the left-hand sides of (22_1)-(22_3) are majorized by $|\Delta\phi| + 2|h|/m$. Since the elements of the matrix $(A_i \cdot A_j)$ are majorized by $|A_i||A_j| < 2 \cdot 2 = 4$, it follows that the two rowed minors are majorized by 32. Hence, if (20) holds,

$$|\alpha_i| \leqq 96(|\Delta\phi| + 2|h|/m)\det^{-2}(A_1, A_2, A_3),$$

since $\det^2(A_1, A_2, A_3) = \det(A_i \cdot A_j)$. Hence, (18) implies

$$(24) \qquad |\Delta X_u| \det^2(A_1, A_2, A_3) \leqq C(|\Delta\phi| + 2|h|/m),$$

if C is sufficiently large, say $C = 576$. Hence, by (16) and (17),

$$(25) \qquad |\Delta X_u| \det^2(A_1, A_2, A_3) \leqq C(\epsilon + 2\delta/m) \leqq 2C\epsilon$$

if $|h| < \delta$.

It will now be shown that

$$(26) \qquad |\Delta X_u| < m \text{ if } |h| < \delta$$

for all (u, v), $(u + h, v)$ in R. By the definition (19) of A_1, A_2, it is seen that the vector product (A_1, A_2) is $N \sin\phi + \frac{1}{2}(\Delta X_u, X_v)$; so that $\det(A_1, A_2, A_3)$ is $\sin\phi + \frac{1}{2}\det(\Delta X_u, X_v, N)$. Consequently,

$$(27) \qquad |\det(A_1, A_2, A_3)| \geqq m - \tfrac{1}{2}|\Delta X_u|.$$

Suppose, if possible, that the first inequality in (26) is violated for a pair of points $(u + h, v)$, (u, v) of R, where $|h| < \delta$. Clearly, it can then be supposed that the first inequality sign in (26) can be replaced by equality if $(u + h, v)$, (u, v) are suitable chosen. By (27), $|\det(A_1, A_2, A_3)| \geqq \frac{1}{2}m > 0$; and so, (25) implies $|\Delta X_u| \leqq (2C\epsilon)/(\frac{1}{2}m)^2$. By (15), this means $|\Delta X_u| < m$, which contradicts the assumption $|\Delta X_u| = m$. Consequently, (26) holds.

Hence, (25) and (27) imply (18) for all $(u + h, v)$, (u, v) in R. This completes the proof of the existence statement in (*).

6. There remains to prove the uniqueness statement contained in the assertion (*). Let $S_1 : z = z_1(x_1, y_1)$ and $S_2 : z = z_2(x_2, y_2)$ be two surfaces satisfying the statement of the first paragraph of § 1, where the function $\phi = \phi(u, v)$ for both surfaces is the given ϕ. It will be verified that S_1 and S_2 are identical, up to a Euclidean movement.

In the (x_1, y_1)-parametrization of S_1, the squared element of arc-length on S_1 is

$$(28) \qquad ds^2 = (1 + p^2)dx_1^2 + 2pq\,dx_1dy_1 + (1 + q^2)dy_1^2,$$

where $p = \partial z_1/\partial x_1$ and $q = \partial z_1/\partial y_1$. The surfaces S_1, S_2 have parametrizations $X = X_1(u, v)$, $X = X_2(u, v)$ of class C', in which (2) holds with the same given ϕ. Let $x = x_2(x_1, y_1)$, $y = y_2(x_1, y_1)$ be the result of the transformations $(x_2, y_2) \to (u, v) \to (x_1, y_1)$. Thus S_2 has a parametrization of the form

$$(29) \qquad S_2 : x = x_2(x_1, y_1), y = y_2(x_1, y_1), z = z_2(x_2(x_1, y_1), y_2(x_1, y_1)),$$

which is of class C'. In the (x_1, y_1)-parametrization, it follows from the standard transformation rule for ds^2, that (28) holds on S_2. Since the coefficients in (28) are of class C_1, it follows from (**) in Part II below that the transformation $(x_1, y_1) \rightarrow (x_2, y_2)$ is of class C''; and so the (x_1, y_1)-parametrization of both S_1 and S_2 are of class C''.

It can easily be verified that, if the second fundamental form in the (u, v)-parameters are calculated formally by the standard transformation rule (or equivalently by, $h_{11} = -X_u \cdot N_u$, $h_{12} = h_{21} = -X_u \cdot N_v$, $h_{22} = -X_v \cdot N_v$) for both S_1 and S_2, then $h_{ik}du^i du^k = \pm 2 \sin \phi \, dudv$; and it can, therefore, be supposed that (8) holds. Hence, in terms of the (x_1, y_1)-parameters, S_1 and S_2 have the same second fundamental form.

Thus, the surfaces $S_1 : z = z_1(x_1, y_1)$ and (29) are of class C'' and have the same first and second fundamental forms. It follows from the uniqueness theorem in [3], § 2 that S_1 and S_2 are identical, up to a Euclidean movement. This completes the proof of the assertion (*).

Part II.

7. In this part, the following theorem will be proved:

(**) *Let* $(g_{ik}) = (g_{ik}(u^1, u^2))$, *where* $i, k = 1, 2$, *be a positive-definite symmetric matrix of class* C' *in a vicinity of* $(u^1, u^2) = (0, 0)$. *Let*

$$u^i = u^i(U^1, U^2), \text{ where } i = 1, 2, \qquad (u^i(0,0) = 0), \tag{30}$$

be a transformation of class C' *in a vicinity of* $(U^1, U^2) = (0, 0)$ *with a non-vanishing Jacobian carrying*

$$ds^2 = g_{ik} du^i du^k \tag{31}$$

into

$$ds^2 = G_{ik} dU^i dU^k. \tag{32}$$

If (30) *has the property that the symmetric matrix* $(G_{ik}) = (G_{ik}(U^1, U^2))$ *is of class* C', *then the transformation* (30) *is of class* C''.

This theorem is a generalization of the lemma of [4], § 12. In addition to the application of (**) in § 6, above, several other consequences of (**) can be mentioned:

(i) From Wintner's discussion [9] of the notion of isometry in differential geometry, the usefulness of (**) is at once apparent. In fact, (**) and its analogues (cf. the remark at the end of this section) show that if two surfaces of class C^n, where $n \geqq 1$, are isometric (by virtue of a transformation of class C^1), then they are isometric by virtue of a transformation of class C^n. This answers the question raised by Wintner [9], end of § 4, concerning the interpretation of certain standard theorems in the theory of surfaces.

(ii) If the coefficients g_{ik} of (31) are of class C' and (31) possesses a curvature $K = K(u, v)$ in the sense of Weyl [7] (cf., [6]) and K is of class C', then it follows from [1], § 6 and from (**) that it is possible to introduce local geodesic parallel coordinates (U^1, U^2) and that the transformation (30) is of class C''. (For contrast, it can be mentioned that such a transformation (30) need not exist if a bounded curvature K does not exist (cf. [1], § 2); the transformation (30) exists and is of class C' when a bounded curvature K does exist (cf. the proofs of Theorems 1 and 2 in [1] and (III) in [5]), but need not be of class C'' if K is not of class C' (cf. [2]).

(iii) If the surface S has a parametrization of class C'' in terms of some parameters, say (u^1, u^2), and if (30) is a transformation of class C' such that the first fundamental form $G_{ik}dU^i dU^k$ and formal (cf. § 6 above) second fundamental form $H_{ik}dU^i dU^k$ are of class C', then, by (**) and the considerations of [4], § 14, the surface S is of class C''' in a suitable parametrization, say $z = z(x, y)$.

To illustrate the principle (iii), let S possess a negative curvature of class C', then it is possible to introduce (locally) the asymptotic lines as coordinate curves $U^1 = \text{const.}$, $U^2 = \text{Const.}$, and the transformation (30) is of class C', [4], § 6. If the coefficients in the resulting squared element of arc-length (32) are of class C', then the above principle is applicable, since $H_{11} = H_{22} = 0$ and $H_{12}{}^2 = -\det(G_{ik})/K > 0$ are of class C'. Thus, the asymptotic line parametrization of a surface (of negative curvature K of class C') cannot be of class C'' unless the surface has some parametrization of class C'''. The case $K \equiv -1$ of this illustration is proved in [4] by using the lemma, [4], § 12, mentioned above.

As another illustration of (iii), let S possess distinct principle curvatures of class C', then it is possible to introduce (locally) the lines of curvature as coordinate curves $U^1 = \text{const.}$, $U^2 = \text{Const.}$, and the transformation (30) is of class C', [4], § 17. If the coefficients in the resulting squared element of arc-length (32) are of class C', then the above principle is applicable, since $H_{12} = 0$ and H_{11}/G_{11}, H_{22}/G_{22}, being the respective principal curvatures, are of class C'. Thus, the line of curvatures parametrization of a surface (having distinct principal curvatures of class C') cannot be of class C'' unless the surface has some parametrization of class C'''.

Whether or not (**) is true if the symmetric matrix (g_{ik}) is either n by n and non-singular (instead of being 2 by 2 and positive-definite) will remain undecided. (The proof of (**) and this generalization would follow, for instance, if it could be shown that the transformation (30) can be approximated by smooth transformations in such a way that the Christoffel symbols belonging to the forms corresponding to (32) tend uniformly to those belonging to (32)). An indication of the truth in the n by n, positive-definite case is given by the results in [8].

The theorem (**) is obviously false if (g_{ik}) is singular. For consider the cases $u = U$, $v = v(U, V)$ and $ds^2 = du^2 + 0 \cdot dudv + 0 \cdot dv^2$ of (30) and (31), respectively. The corresponding form (32) is $ds^2 = dU^2 + 0 \cdot dUdV + 0 \cdot dV^2$, while $v(U, V)$ need not be of class C''.

The replacement of C', C'' by C^m, C^{m+1}, respectively, where $m > 1$, in (**) leads to a correct but easily proved theorem, in view of the transformation rule for Christoffel symbols. The difficulty in the proof of (**) is to establish the validity of this rule when $m = 1$.

8. *Proof of* (**). Since the coefficient functions of both (31) and (32) are of class C', the Christoffel symbols of the second kind, $\gamma_{jk}{}^i = \gamma_{jk}{}^i(u^1, u^2)$ and $\Gamma_{jk}{}^i = \Gamma_{jk}{}^i(U^1, U^2)$, belonging to (31) and (32), respectively, exist and are continuous. The two systems of differential equations

$$u^{i\prime\prime} + \gamma_{jk}{}^i u^{j\prime} u^{k\prime} = 0, \text{ where } i = 1, 2, \qquad (' = d/ds), \tag{33}$$

and

$$U^{i\prime\prime} + \Gamma_{jk}{}^i U^{j\prime} U^{k\prime} = 0, \text{ where } i = 1, 2, \qquad (' = d/ds), \tag{34}$$

define the geodesics belonging to (31) and (32), respectively. Formally, it cannot, however, be verified that the solutions of (34) correspond, by virtue of (30), to solutions of (33), and conversely. (Of course, such a verification could be made by considering the geodesics as extremals (but not "minimizing" curves, cf. (II) in [5]) of a calculus of variations problem.)

If a point (u^1, u^2) and/or (U^1, U^2) is sufficiently near $(0, 0)$, then there exist arcs minimizing the distance, in the metric (31) and/or (32), between $(0, 0)$ and that point (Hilbert). All such arcs are solutions of (33) and/or (34); cf. the proof of (I) in [5]. If $u^i = u^i(s)$ and $U^i = U^i(s)$ is such a geodesic, where the initial conditions are chosen so that (30) gives

$$u^{i\prime} = u_j{}^i U^{j\prime}, \text{ where } u_j{}^i = \partial u^i / \partial U^j, \tag{35}$$

then both $u^i(s)$ and $U^i(s)$ are of class C''. (Incidentally, not all geodesics minimize the distance between sufficiently close points on them; cf. (II) in [5]).

Corresponding to a geodesic $u^i = u^i(s)$, $U^i = U^i(s)$, related by (30) and (35), consider the linear differential equations

$$w^{i\prime} + \gamma_{jk}{}^i w^j u^{k\prime} = 0, \text{ where } i = 1, 2 \tag{36}$$

and

$$W^{i\prime} + \Gamma_{jk}{}^i W^j U^{k\prime} = 0, \text{ where } i = 1, 2, \tag{37}$$

with $\gamma_{jk}{}^i = \gamma_{jk}{}^i(u^1(s), u^2(s))$ and $\Gamma_{jk}{}^i = \Gamma_{jk}{}^i(U^1(s), U^2(s))$ in (36) and (37), respectively. A solution $W^i = W^i(s)$ of (37) is a field of parallel vectors along the geodesic $U^i = U^i(s)$. Hence, the scalar products $G_{ik}W^iW^k$, $G_{ik}W^iU^{k\prime}$, which are first integrals of (37), are independent of s. But scalar products are invariant under transformations of class C'; and so

$$w^i = u_j{}^i W^j, \text{ where } i = 1, 2, \tag{38}$$

with $u_j{}^i = u_j{}^i(U^1(s), U^2(s))$, is a solution of (36) and, in particular, is of class C'.

It is clear from (34) and the fact that $G_{ik}U^{i\prime}U^{k\prime} = 1$ along solutions of (34), while (32) is positive-definite, that $| U^{i\prime\prime} | \leqq$ Const., whenever $(U^1(s), U^2(s))$ is sufficiently near $(0, 0)$, where Const. is independent of the geodesic $U^i = U^i(s)$. This implies that if J_Δ is a geodesic arc $U^i = U^i(s)$ $= U^i(s, \Delta U^1)$ joining $(0, 0)$ and $(\Delta U^1, 0)$, where $| \Delta U^1 | \neq 0$ is sufficiently small, then $(U^{1\prime}(s), U^{2\prime}(s))$ tends, uniformly on J_Δ, to $(G_{11}{}^{-\frac{1}{2}}(0, 0), 0)$, as $\Delta U^1 \to 0$. Similarly, if $W^i = W^i(s) = W^i(s, \Delta U^1)$ is a field of parallel vectors on J_Δ determined by initial conditions $(W_0{}^1, W_0{}^2)$ at $(0, 0)$, which are independent of ΔU^1, then $(W^1(s), W^2(s))$ tends, uniformly on J_Δ, to $(W_0{}^1, W_0{}^2)$, as $\Delta U \to 0$. Finally, if Δs is the length of J_Δ, then $\Delta s/\Delta U^1$ $\to G_{11}{}^{\frac{1}{2}}(0, 0)$, as $\Delta U^1 \to 0$.

Let the geodesic arc J_Δ be chosen so that $u^i = u^i(s)$, determined from (30), is a solution of (33). Let f_0, where f is $w^i, u_j{}^i, \cdots$, denote the value of f at $(0, 0)$ and let Δf denote the difference between the values of f at the points $(\Delta U^1, 0)$ and $(0, 0)$. Then, by (38),

$$\Delta w^i = (\Delta u_j{}^i) W_0{}^j + u_j{}^i \Delta W^j. \tag{39}$$

By (37) and the mean-value theorem of differential calculus,

$$\Delta W^j = - (\Gamma_{nm}{}^j W^n U^{m\prime}) \Delta s,$$

where the coefficient of Δs is evaluated at some intermediary point on J_Δ (depending on j and ΔU^1). By the remarks of the last paragraph, $\Delta W^j/\Delta U^1$ tends to $- \Gamma_{n10}{}^j W_0{}^n$, as $\Delta U^1 \to 0$. Similarly, $\Delta w^i/\Delta U^1$ tends, as $\Delta U^1 \to 0$, to $- \gamma_{kmo}{}^i w_0{}^k u_0{}^{m\prime} G_{110}{}^{-\frac{1}{2}}$. By (35) and (38), the latter limit is $- \gamma_{kmo}{}^i u_{no}{}^k u_{10}{}^m W_0{}^n$.

Hence (39) implies that $(\Delta u_j^i/\Delta U^1) W_0^j$ tend to a limit, as $\Delta U^1 \to 0$. Since W_0^j is arbitrary, it follows that, as $\Delta U^1 \to 0$, the quotient

$$\Delta u_j^i/\Delta U^1 = \{u_j^i(\Delta U^1, 0) - u_j^i(0,0)\}/\Delta U^1$$

tends to the limit $\Gamma_{j1}{}^n u_n{}^i - \gamma_{km}{}^i u_j{}^k u_1{}^m$, evaluated at $(0,0)$. Since ΔU^1 can be replaced by ΔU^2 in this argument, and $(0,0)$ by any point (U^1, U^2) sufficiently near $(0,0)$, it follows that the second order partial derivatives of the functions in (30) exist and can be calculated from the standard transformation rule for Christoffel symbols of the second kind. Consequently, (**) is proved.

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "On the local uniqueness of geodesics," *American Journal of Mathematics*, vol. 72 (1950), pp. 723-730.

[2] ———, "On geodesic coordinates," *ibid.*, vol. 73 (1951), pp, 949-954.

[3] ——— and A. Wintner, "On the fundamental equations of differential geometry," *ibid.*, vol. 72 (1950), pp. 757-774.

[4] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[5] ——— and A. Wintner, "On the problems of geodesics in the small," *ibid.*, vol. 73 (1951), pp. 132-148.

[6] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *ibid.*, vol. 60 (1938), pp. 129-138.

[7] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

[8] A. Wintner, "On Riemann metrics of constant curvature," *American Journal of Mathematics*, vol. 73 (1951), pp. 569-575.

[9] ———, "On isometric surfaces," *ibid.*, vol. 74 (1952), pp. 198-214.

ON THE KINEMATIC FORMULA IN THE EUCLIDEAN SPACE OF N DIMENSIONS.*

By Shiing-shen Chern.

Introduction. The idea of considering the kinematic density in problems of geometrical probability was originated by Poincaré. It was further exploited by L. A. Santalò and W. Blaschke in their work on integral geometry [1], culminating in the following theorem:

Let Σ_0, Σ_1 be two closed surfaces in space, which are twice differentiable, and let D_0, D_1 be the domains bounded by them. Let V_i, $\chi_i = K_i/4\pi$ be the volume and Euler characteristic of D_i and let A_i, M_i be the area and the integral of mean curvature of Σ_i, $i = 0, 1$. Suppose Σ_0 fixed and Σ_1 moving. Then the integral of $K(D_0 \cdot D_1) = 4\pi\chi(D_0 \cdot D_1)$ over the kinematic density of Σ_1 is given by the formula

$$(1) \qquad \int K(D_0 \cdot D_1)\dot{\Sigma}_1 = 8\pi^2(V_0K_1 + A_0M_1 + M_0A_1 + K_0V_1).$$

This formula includes most formulas in Euclidean integral geometry as special or limiting cases. The purpose of this paper is to apply E. Cartan's method of moving frames and to derive the generalization of this formula in an Euclidean space of n dimensions. By doing this, we hope that some insight can be gained on integral geometry in a general homogeneous space. Moreover, one of the ideas introduced, the consideration of measures in spaces which are now called fiber bundles, will most likely find further applications. The main procedures of our proof have been given in a previous note [2].

We consider a compact orientable hypersurface Σ, twice differentiably imbedded in an Euclidean space E of n ($\geqq 2$) dimensions. At a point P of Σ there are $n-1$ principal curvatures κ_α, $\alpha = 1, \cdots, n-1$, whose i-th elementary symmetric function we shall denote by S_i, $i = 0, \cdots, n-1$, where $S_0 = 1$ by definition. Let dA be the element of area of Σ, and let

$$(2) \qquad M_i = \int_\Sigma S_i dA \Big/ \binom{n-1}{i}, \qquad i = 0, 1, \cdots, n-1.$$

These M_i are integro-differential invariants of Σ. In particular, M_0 is the

* Received May 10, 1951.

area and M_{n-1} is a numerical multiple of the degree of mapping of Σ into the unit hypersphere defined by the field of normals.

Take now two such hypersurfaces Σ_0, Σ_1, whose invariants we distinguish by superscripts. The volume of the domain D_i bounded by Σ_i we denote by V_i, $i = 0, 1$. Let Σ_0 be fixed and Σ_1 be moving, and let $\dot{\Sigma}_1$ be the kinematic density of Σ_1. We suppose our hypersurfaces to be such that for all positions of Σ_1 the intersection $D_0 \cdot D_1$ has a finite number of components. Then the Euler-Poincaré characteristic $\chi(D_0 \cdot D_1)$ is well defined. If I_{n-1} denotes the area of the unit hypersphere in E and if

$$J_n = I_1 I_2 \cdots I_{n-1}, \tag{3}$$

the *kinematic formula* in E is

$$\begin{aligned} \int K(D_0 \cdot D_1)\dot{\Sigma}_1 \\ = J_n\{M_{n-1}{}^{(0)}V_1 + M_{n-1}{}^{(1)}V_0 + \frac{1}{n}\sum_{k=0}^{n-2}\binom{n}{k+1} M_k{}^{(0)} M_{n-2-k}{}^{(1)}\}, \end{aligned} \tag{4}$$

where

$$K(D_0 \cdot D_1) = I_{n-1}\chi(D_0 \cdot D_1). \tag{5}$$

For $n = 3$ this reduces to the formula (1). The formula for $n = 4$ is

$$\begin{aligned} \int K(D_0 \cdot D_1)\dot{\Sigma}_1 \\ = 16\pi^4(M_3{}^{(0)}V_1 + M_3{}^{(1)}V_0 + M_0{}^{(0)}M_2{}^{(1)} + M_0{}^{(1)}M_2{}^{(0)} + \tfrac{3}{2}M_1{}^{(0)}M_1{}^{(1)}). \end{aligned} \tag{6}$$

1. Measures in spaces associated with a Riemann manifold. We shall first review a few notions in Riemannian geometry, in a form which will be useful for our later purpose.

Let M be an orientable Riemann manifold of class $\geqq 3$ and dimension n. Associated with M are the spaces B_h $(h = 1, \cdots, n)$ formed by the elements $Pe_1 \cdots e_h$, each of which consists of a point P of M and an ordered set of h mutually perpendicular tangent unit vectors $e_1, \cdots, e_h$ at P. When $h = n$, such an element will be called a frame. In the current terminology B_n is a principal fiber bundle over M with the rotation group as structural group and B_h are the associated bundles [3]. We shall introduce a measure in B_h. Since B_h is clearly an orientable differentiable manifold, this can be done by defining an exterior differential form of degree $\frac{1}{2}(h+1)(2n-h)(=\dim$ of $B_h)$.

There is a natural mapping $\psi_h : B_n \to B_h$ defined by taking as the image of $Pe_1 \cdots e_n$ the element $Pe_1 \cdots e_h$. It induces a dual homomorphism of the differential forms of B_h into those of B_n. This process has in a sense a converse. In fact, let

$$e_r = \sum_s u_{rs} e^*_s, \qquad h+1 \leqq r, s \leqq n \tag{7}$$

be a rotation of the last $n-h$ vectors. A differential form of B_n which is invariant under the action of (7) can be regarded as a form of B_n.

The well-known parallelism of Levi-Civita can be interpreted as defining a set of $n(n+1)/2$ linearly independent Pfaffian forms in B_n, which we shall denote by ω_i, $\omega_{ij}(=-\omega_{ji})$, $1 \leqq i, j \leqq n$. To give it a brief description [4] we start from the following useful lemma on exterior forms: *Let ω_i be linearly independent Pfaffiian forms, and let $\pi_{ij}=-\pi_{ji}$ be Pfaffian forms such that* [1]

$$\sum_j \omega_j \wedge \pi_{ji} = 0. \tag{8}$$

Then $\pi_{ij}=0$. In fact, it follows from (8) that

$$\pi_{ji} = \sum_{k=1}^{n} a_{jik}\omega_k.$$

Then a_{jik} is skew-symmetric in its first two indices, because the π_{ji} are, and is symmetric in its last two indices, on account of (8). Therefore $a_{jik}=0$ or $\pi_{ji}=0$.

For geometric reasons we denote by dP the identity mapping in the tangent space at P, which maps every tangent vector into itself. Then dP can be written in the form

$$dP = \sum_i \omega_i \otimes \mathfrak{e}_i, \tag{9}$$

where the multiplication is tensor product, and the ω_i are Pfaffian forms in B_n and are linearly independent. The fundamental theorem on local Riemannian geometry asserts that there exists a uniquely determined set of Pfaffian forms ω_i, ω_{ij} in B_n, linearly independent, which satisfy (9) and

$$d\omega_i = \sum_j \omega_j \wedge \omega_{ji}. \tag{10}$$

In fact, the uniqueness follows from the above lemma.

For our purpose we shall study the effect of the rotation (7) on these forms. Denote the new forms by the same symbols with asterisks. Clearly we have

$$\omega_\alpha^* = \omega_\alpha, \qquad \omega_r^* = \sum_s u_{sr}\omega_s, \qquad 1 \leqq \alpha \leqq h,\ h+1 \leqq r, s \leqq n. \tag{11}$$

Taking the exterior derivatives of both sides of these equations and making use of (10), we get

---

[1] We shall, following Bourbaki, use wedge product to denote exterior multiplication. It will sometimes be dropped, when the meaning is clear. Parentheses will be used to denote ordinary products of differential forms.

$$\begin{aligned}(12)\qquad &\sum_\beta \omega_\beta \wedge (\omega_{\beta\alpha}{}^* - \omega_{\beta\alpha}) + \sum_r \omega_r{}^* \wedge (\omega_{r\alpha}{}^* - \sum_s u_{sr}\omega_{s\alpha}) = 0,\\ &\sum_\alpha \omega_\alpha \wedge (\omega_{\alpha r}{}^* - \sum_s u_{sr}\omega_{\alpha s}) + \sum_s \omega_s{}^* \wedge \phi_{sr}{}^* = 0,\end{aligned}$$

where $\phi_{sr}{}^*$ are Pfaffian forms skew-symmetric in the indices s, r. The system of equations (12) is of the same form as (8), and the above lemma is then applicable. It follows that

$$(13)\qquad \omega_{\beta\alpha}{}^* = \omega_{\beta\alpha}, \qquad \omega_{\alpha r}{}^* = \sum_s u_{sr}\omega_{\alpha s}.$$

If we put

$$(14)\qquad \Omega_\alpha = \prod_r \omega_{\alpha r},$$

we see from (13) that Ω_α is invariant under the action of (7). The same is therefore true of the form

$$(15)\qquad L_{n,h} = \prod_\alpha \Omega_\alpha \prod_{\alpha<\beta} \omega_{\alpha\beta} \prod_i \omega_i.$$

This form is clearly not identically zero, and we define it to be the density in B_h. It gives rise to a measure in B_h.

2. Differential geometry of a submanifold in Euclidean space. As a further preparation we need some notions on the geometry of a hypersurface in Euclidean space. As no additional complication is involved, we develop them for a submanifold V of p dimensions, which is twice differentiably imbedded in E. We agree in this section on the following ranges of indices:

$$(16)\qquad 1 \leqq \alpha, \beta, \gamma \leqq p, \qquad p+1 \leqq r, s, t \leqq n, \qquad 1 \leqq i, j, k \leqq n.$$

Since E is a Riemann manifold, the discussions of the last section are valid. In this case B_n is naturally homeomorphic to the group of proper motions in E. To study V we consider the submanifold of B_n characterized by the conditions that $P \varepsilon V$ and that the $\mathfrak{e}_\alpha$ are tangent vectors to V at P. If we denote by the same notation the forms on this submanifold induced by the identity mapping, we have

$$(17)\qquad \omega_r = 0.$$

From (10) it follows that

$$d\omega_r = \sum_\alpha \omega_\alpha \wedge \omega_{\alpha r} = 0.$$

Since the ω_α are linearly independent, we have

$$(18)\qquad \omega_{r\alpha} = \sum_\beta A_{r\alpha\beta}\omega_\beta$$

where the $A_{r\alpha\beta}$ are symmetric in α, β:

$$A_{r\alpha\beta} = A_{r\beta\alpha}. \tag{19}$$

From these Pfaffian forms it is possible to construct some significant "ordinary" quadratic differential forms. The first is a set

$$\Phi_r = \sum_{\alpha} (\omega_{r\alpha}\omega_\alpha) = \sum_{\alpha,\beta} A_{r\alpha\beta}(\omega_\alpha\omega_\beta), \tag{20}$$

which generalizes the second fundamental form in ordinary surface theory. The second is

$$\Psi = \sum_{r,\alpha} (\omega_{r\alpha})^2 = \sum_{r,\alpha,\beta,\gamma} A_{r\alpha\beta}A_{r\alpha\gamma}(\omega_\beta\omega_\gamma), \tag{21}$$

generalizing the third fundamental form. The latter seems to deserve some attention. However, so far as the writer is aware, it has not been considered in the literature.

For a hypersurface we have $p = n - 1$, and we shall write Φ, $A_{\alpha\beta}$ for Φ_n, $A_{n\alpha\beta}$ respectively. The $n - 1$ roots of the characteristic equation

$$| A_{\alpha\beta} - \kappa\delta_{\alpha\beta} | = 0 \tag{22}$$

are called the principal curvatures.

In the case of the Euclidean space E we can also write ω_i, ω_{ij} as scalar products, thus:

$$\omega_i = dP \cdot \mathfrak{e}_i, \qquad \omega_{ij} = d\mathfrak{e}_i \cdot \mathfrak{e}_j. \tag{23}$$

3. A formula on densities. The situation we are going to consider consists of two hypersurfaces Σ_0, Σ_1 in E, with Σ_0 fixed and Σ_1 moving, which intersect in a manifold V^{n-2} of dimension $n - 2$, such that at a point of V^{n-2} the normals to Σ_0, Σ_1 never coincide. We denote by ϕ, $\phi \neq 0, \pi$, the angle between these normals and by $\dot{\Sigma}_1$ the kinematic density of Σ_1. An $(n-2)$-frame on V^{n-2} has a density on each of V^{n-2}, Σ_0, Σ_1, to be denoted by L_V, $\dot{L}_0$, L_1 respectively. Our formula to be proved can be written

$$\dot{L}_V\dot{\Sigma}_1 = \sin^{n-1}\phi\dot{L}_0\dot{L}_1\dot{\phi}. \tag{24}$$

Throughout this section we shall agree on the following ranges of indices:

$$1 \leqq i, j, k \leqq n, \qquad 1 \leqq \alpha, \beta \leqq n-2, \qquad 1 \leqq A, B \leqq n-1. \tag{25}$$

Let $O\mathfrak{a}_1 \cdots \mathfrak{a}_n$ be the fixed frame and $O'\mathfrak{a}'_1 \cdots \mathfrak{a}'_n$ the moving frame. For a given relative position between $O\mathfrak{a}_1 \cdots \mathfrak{a}_n$ and $O'\mathfrak{a}'_1 \cdots \mathfrak{a}'_n$ let $P\mathfrak{e}_1 \cdots \mathfrak{e}_{n-2}$ be an $(n-2)$-frame on V^{n-2}. We complement this into a frame $P\mathfrak{e}_1 \cdots \mathfrak{e}_n$ such that $\mathfrak{e}_n$ is normal to Σ_0 and also into a frame $P'\mathfrak{e}'_1 \cdots \mathfrak{e}'_n$

such that e'_n is normal to Σ_1 at P, and $P' = P$, $e'_\alpha = e_\alpha$. Between e_{n-1}, e_n, e'_{n-1}, e'_n we have then the relations

$$e'_{n-1} = \cos\phi e_{n-1} + \sin\phi e_n, \qquad e'_n = -\sin\phi e_{n-1} + \cos\phi e_n. \tag{26}$$

From this we derive the following useful relation

$$de'_{n-1} \cdot e'_n = d\phi + de_{n-1} \cdot e_n. \tag{27}$$

Let us now express the relations between the frames so introduced by the equations

$$\begin{aligned} P &= O + \sum_i x_i a_i, \qquad & e_i &= \sum_k u_{ik} a_k, \\ P' &= O' + \sum_i x'_i a'_i, \qquad & e'_i &= \sum_k u'_{ik} a'_k. \end{aligned} \tag{28}$$

We shall denote the differentiation by d' when $O'a'_1 \cdots a'_n$ is regarded as fixed. In other words, d' is differentiation relative to the moving frame. Then we have, from (28),

$$dO' = dP - d'P - \sum_i x'_i da'_i. \tag{29}$$

It follows that, on neglecting terms in da'_i,

$$\begin{aligned} &\prod_\alpha (dP \cdot e_\alpha) \prod_i (dO' \cdot a'_i) \equiv \prod_\alpha (dP \cdot e_\alpha) \prod_i (dP \cdot a'_i - d'P \cdot a'_i) \\ &\equiv \prod_\alpha (dP \cdot e_\alpha) \prod_i (dP \cdot e'_i - d'P \cdot e'_i) \\ &\equiv \pm \prod_\alpha (dP \cdot e_\alpha)(d'P \cdot e'_\alpha)(dP \cdot e'_{n-1} - d'P \cdot e'_{n-1})(dP \cdot e'_n - d'P \cdot e'_n) \\ &\equiv \pm \sin\phi \prod_A (dP \cdot e_A)(d'P \cdot e'_A). \end{aligned} \tag{30}$$

These are to be taken as congruences mod da'_i. In particular, the last step follows from the fact that e'_n is normal to Σ_1 at P and that the product of n factors involving dP is zero, because the locus of P is a hypersurface Σ_0.

In order to get a further reduction of the left-hand side of (24) we start from the formula

$$de'_i - d'e'_i = \sum_k u'_{ik} da'_k. \tag{31}$$

From the invariance of the kinematic density under a rotation it follows that

$$\prod_{i<j} (da'_i \cdot a'_j) = \prod_{i<j} ((de'_i - d'e'_i) \cdot e'_j).$$

Then we have

$$
\begin{aligned}
\prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{i<j}(da'_i\cdot a'_j) &= \prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{i<j}(de'_i - d'e'_i)\cdot e'_j \\
&= \prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{\alpha<i}((de_\alpha - d'e_\alpha)\cdot e'_i)\,((de'_{n-1} - d'e'_{n-1})\cdot e'_n) \\
&= \prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)\prod_{\alpha<\beta}((de_\alpha - d'e_\alpha)\cdot e'_\beta) \\
&\qquad \wedge \prod_\alpha\{(de_\alpha\cdot e'_{n-1} - d'e_\alpha\cdot e'_{n-1})(de_\alpha\cdot e'_n - d'e_\alpha\cdot e'_n)\} \\
&\qquad \wedge (de'_{n-1}\cdot e'_n - d'e'_{n-1}\cdot e'_n) \\
&\equiv \pm \prod_{\alpha<\beta}(de_\alpha\cdot e_\beta)(d'e_\alpha\cdot e_\beta)\prod_\alpha(de_\alpha\cdot e'_{n-1} - d'e_\alpha\cdot e'_{n-1}) \\
&\qquad \wedge (de_\alpha\cdot e'_n)(de'_{n-1}\cdot e'_n) \\
&\equiv \pm \sin^{n-2}\phi \prod_{\alpha<A}\{(de_\alpha\cdot e_A)(d'e_\alpha\cdot e'_A)\}d\phi.
\end{aligned}
\tag{32}
$$

Here the congruences are to be understood mod $dP\cdot e_A$, $d'P\cdot e'_A$. The step next to the last follows from the relations

$$d'e'_A\cdot e'_n = -d'e'_n\cdot e'_A \equiv 0, \quad \mathrm{mod}\ d'P\cdot e'_A,$$

which in turn are consequences of (18). In the reduction of the last step we make use of the relations (26), (27), and

$$de_A\cdot e_n = -de_n\cdot e_A \equiv 0, \quad \mathrm{mod}\ dP\cdot e_A.$$

If we notice that

$$\dot{\Sigma} = \prod_i(dO'\cdot a'_i)\prod_{i<j}(da'_i\cdot a'_j),$$

and recall the expressions for $\dot{L}_V$, $\dot{L}_0$, $\dot{L}_1$, then (30) and (32) together give the formula (24).

4. Total curvature and Euler characteristic. The success of our procedure depends on the possibility of expressing the Euler-Poincaré characteristic of a domain bounded by a hypersurface Σ by an integral over Σ, a result known as the Gauss-Bonnet formula. Let Λ be the volume element of the unit hypersphere in E, and N^+ the field of outward normals of Σ. By means of N^+ we define the normal mapping of Σ. The Gauss-Bonnet formula in this particular case can be written

$$\int_{N^+}\Lambda = \chi(D)I_{n-1}, \tag{33}$$

where D is the domain bounded by Σ, and $\chi(D)$ is its Euler-Poincaré characteristic. The left-hand side of this equation is sometimes called the total curvature of the domain.

In our later application the domain D will not be bounded by a smooth hypersurface but will be such that its boundary consists of a finite number of hypersurfaces which intersect in a number of submanifolds V^{n-2} of dimension $n-2$. To the integral of Λ over the outward normals we must then add the integral over the vectors belonging to the angle subtended by the outward normals of the two hypersurfaces. To express the latter analytically let us use the notation of the last section, together with the ranges of indices (25). In addition we denote by $\mathfrak{v}_A$, $\mathfrak{v}'_A$ the unit vectors in the principal directions of Σ_0, Σ_1 respectively. For a differentiation on Σ_0 we can then write

$$\theta_A = dP \cdot \mathfrak{v}_A, \qquad \kappa_A \theta_A = d\mathfrak{e}_n \cdot \mathfrak{v}_A, \tag{34}$$

where κ_A are the principal curvatures. Similarly, for a differentiation on Σ_1 we have

$$\theta'_A = dP \cdot \mathfrak{v}'_A, \qquad \kappa'_A \theta'_A = d\mathfrak{e}'_n \cdot \mathfrak{v}'_A, \tag{35}$$

κ'_A being the principal curvatures of Σ_1. Since the $\mathfrak{e}_\alpha$ lie in the intersection of the tangent hyperplanes, we have relations of the form

$$\mathfrak{e}_\alpha = \sum_A c_{\alpha A} \mathfrak{v}_A = \sum_A c'_{\alpha A} \mathfrak{v}'_A \tag{36}$$

To simplify notation we introduce the unit vectors $\mathfrak{v}$, $\mathfrak{w}$ in the directions of the angle bisectors of $\mathfrak{e}_n$, $\mathfrak{e}'_n$. Then we have

$$\mathfrak{e}_n = (\cos \tfrac{1}{2}\phi)\mathfrak{v} - (\sin \tfrac{1}{2}\phi)\mathfrak{w}, \qquad \mathfrak{e}'_n = (\cos \tfrac{1}{2}\phi)\mathfrak{v} + (\sin \tfrac{1}{2}\phi)\mathfrak{w}, \tag{37}$$

or

$$\mathfrak{e}_n + \mathfrak{e}'_n = 2(\cos \tfrac{1}{2}\phi)\mathfrak{v}, \qquad -\mathfrak{e}_n + \mathfrak{e}'_n = 2(\sin \tfrac{1}{2}\phi)\mathfrak{w}. \tag{38}$$

Let $\mathfrak{x}$ be a unit vector between $\mathfrak{e}_n$. $\mathfrak{e}'_n$, and $\mathfrak{y}$ the unit vector perpendicular to $\mathfrak{x}$ and in the plane of $\mathfrak{e}_n$, $\mathfrak{e}'_n$. We can then write

$$\mathfrak{x} = \cos \sigma \mathfrak{v} + \sin \sigma \mathfrak{w}, \qquad \mathfrak{y} = -\sin \sigma \mathfrak{v} + \cos \sigma \mathfrak{w}; \tag{39}$$

$$-\tfrac{1}{2}\phi \leqq \sigma \leqq \tfrac{1}{2}\phi.$$

It follows that the total curvature, i. e., I_{n-1} times the Euler-Poincaré characteristic of D, is given by

$$K = \int_{N^+} \Lambda + \int_{-\frac{1}{2}\phi}^{\frac{1}{2}\phi} d\sigma \int_V \prod_\alpha \{\cos \sigma (d\mathfrak{v} \cdot \mathfrak{e}_\alpha) + \sin \sigma (d\mathfrak{w} \cdot \mathfrak{e}_\alpha)\}. \tag{40}$$

The product in the second integral admits some further simplification. In fact, using (38), we have

$$\prod_\alpha \{\cos \sigma (d\mathfrak{v} \cdot \mathfrak{e}_\alpha) + \sin \sigma (d\mathfrak{w} \cdot \mathfrak{e}_\alpha)\}$$
$$= \prod_\alpha \{\sin(\tfrac{1}{2}\phi - \sigma)(d\mathfrak{e}_n \cdot \mathfrak{e}_\alpha) + \sin(\tfrac{1}{2}\phi + \sigma)(d\mathfrak{e}'_n \cdot \mathfrak{e}_\alpha)\}/\sin^{n-2} \phi.$$

By (36), we get

$$de_n \cdot e_\alpha = \sum_A \kappa_A c_{\alpha A}\theta_A = \sum_{A,\beta} \kappa_A c_{\alpha A} c_{\beta A}(dP \cdot e_\beta),$$

$$de'_n \cdot e_\alpha = \sum_{A,\beta} \kappa'_A c'_{\alpha A} c'_{\beta A}(dP \cdot e_\beta).$$

It follows that

$$\prod_\alpha \{\cos\sigma(d\mathfrak{v} \cdot e_\alpha) + \sin\sigma(d\mathfrak{w} \cdot e_\alpha)\} = D\dot{V}/\sin^{n-2}\phi, \tag{41}$$

where $\dot{V}$ is the volume element of V^{n-2}, and where

$$D = |\sin(\tfrac{1}{2}\phi - \sigma)\sum_A \kappa_A c_{\alpha A} c_{\beta A} + \sin(\tfrac{1}{2}\phi + \sigma)\sum_A \kappa'_A c'_{\alpha A} c'_{\beta A}|. \tag{42}$$

The determinant D can be expanded in the form

$$D = \sum_{p=0}^{n-2} H_p \sin^{n-2-p}(\tfrac{1}{2}\phi - \sigma)\sin^p(\tfrac{1}{2}\phi + \sigma), \tag{43}$$

where

$$H_p = \Sigma \begin{vmatrix} c_{1A_1} & \cdots & c_{1A_q} & c'_{1B_1} & \cdots & c'_{1B_p} \\ & \cdots & & & \cdots & \\ c_{n-2,A_1} & \cdots & c_{n-2,A_q} & c'_{n-2,B_1} & \cdots & c'_{n-2,B_p} \end{vmatrix}^2 \kappa_{A_1}\cdots\kappa_{A_q}\kappa'_{B_1}\cdots\kappa'_{B_p}, \quad p+q=n-2, \tag{44}$$

the summation being extended over all independent combinations $A_1, \cdots, A_q$ and $B_1, \cdots, B_p$ of $1, \cdots, n-1$. To prove this we observe that the expansion of D is of the above form and that the question is only to determine the coefficient of $\kappa_{A_1}\cdots\kappa_{A_q}\kappa'_{B_1}\cdots\kappa'_{B_p}$ in H_p. This coefficient is, up to the factor $\sin^q(\tfrac{1}{2}\phi - \sigma)\sin^p(\tfrac{1}{2}\phi + \sigma)$, the value of D, when we set

$$\kappa_{A_1} = \cdots = \kappa_{A_q} = 1, \qquad \kappa'_{B_1} = \cdots = \kappa'_{B_p} = 1,$$

and equal to zero otherwise. Writing

$$\bar{c}_{\alpha A} = \{\sin(\tfrac{1}{2}\phi - \sigma)\}^{\frac{1}{2}} c_{\alpha A}, \qquad \bar{c}'_{\alpha A} = \{\sin(\tfrac{1}{2}\phi + \sigma)\}^{\frac{1}{2}} c'_{\alpha A},$$

we have

$$D = |\sum_{s=A_1,\ldots,A_q} \bar{c}_{\alpha s}\bar{c}_{\beta s} + \sum_{t=B_1,\ldots,B_p} \bar{c}'_{\alpha t}\bar{c}'_{\beta t}| = \begin{vmatrix} \bar{c}_{1A_1} & \cdots & \bar{c}_{1A_q} & \bar{c}'_{1B_1} & \cdots & \bar{c}'_{1B_p} \\ & \cdots & & & \cdots & \\ \bar{c}_{n-2,A_1} & \cdots & \bar{c}_{n-2,A_q} & \bar{c}'_{n-2,B_1} & \cdots & \bar{c}'_{n-2,B_p} \end{vmatrix}^2.$$

This shows that the coefficient is actually the one asserted in (43), (44).

5. Proof of the kinematic formula. Let Σ_0, Σ_1 be two hypersurfaces twice differentiably imbedded in E, with Σ_0 fixed and Σ_1 moving. We denote by D_i the domain bounded by Σ_i, $i = 0, 1$, and suppose that the intersection $D_0 \cdot D_1$ consists of a finite number of components F_s. The boundary of ΣF_s consists of the sets $\Sigma_1 \cdot D_0$, $\Sigma_0 \cdot D_1$, $\Sigma_0 \cdot \Sigma_1$, so that we can write

$$\int K(D_0 \cdot D_1)\dot{\Sigma}_s = \int K(\Sigma F_s)\dot{\Sigma}_1 \tag{45}$$
$$= \int K(\Sigma_1 \cdot D_0)\dot{\Sigma}_1 + \int K(\Sigma_0 \cdot D_1)\dot{\Sigma}_1 = \int K(\Sigma_0 \cdot \Sigma_1)\dot{\Sigma}_1.$$

The first two integrals are easily evaluated. Take, for instance, the second integral. For every position of Σ_1 the integrand $K(\Sigma_0 \cdot D_1)$ is the integral of Λ over the outward normals to Σ_0 at points of $\Sigma_0 \cdot D_1$. This domain of integration can be decomposed in a different way by first fixing a common point of D_1 and Σ_0, rotating D_1 about this point, and then letting this point vary over D_1 and Σ_0 respectively. The result of this iterated integration is

$$\int K(\Sigma_0 \cdot D_1)\Sigma_1 = J_n K_0 V_1 = J_n M_{n-1}{}^{(0)} V_1. \tag{46}$$

Similarly. using the fact that the kinematic density is invariant under the "inversion" of a motion, we have

$$\int K(\Sigma_1 \cdot D_0)\dot{\Sigma}_1 = J_n K_1 V_0 = J_n M_{n-1}{}^{(1)} V_0. \tag{47}$$

To evaluate the third integral in (45) we use the density formula (24), and the formulas (40)-(44) for the total curvature arising from $\Sigma_0 \cdot \Sigma_1$. We get

$$\int K(\Sigma_0 \cdot \Sigma_1)\dot{\Sigma}_1 = \int (D/\sin^{n-2}\phi)d\sigma \dot{V}\dot{\Sigma}_1 = (1/J_{n-2})\int (D/\sin^{n-2}\phi)d\sigma \dot{L}_V \dot{\Sigma}_1$$
$$= (1/J_{n-2})\int (\sin\phi) D d\sigma d\phi \dot{L}_0 \dot{L}_1$$
$$= b_{n-2}\int H_0 \dot{L}_0 \dot{L}_1 + \cdots + b_0 \int H_{n-2} \dot{L}_0 \dot{L}_1$$
$$= a_{n-2} M_{n-2}{}^{(0)} M_0{}^{(1)} + \cdots + a_0 M_0{}^{(0)} M_{n-2}{}^{(1)}$$

where the a's and b's are numerical constants. These constants can be determined if we take Σ_0, Σ_1 to be two hyperspheres of radii 1 and h respectively. This completes the proof of the kinematic formula.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] W. Blaschke, *Integralgeometrie*, Hamburg, 1936 and 1937.

[2] S. Chern and C. T. Yen, "Sulla formula principale cinematica dello spazio ad n dimensioni," *Bolletino della Unione Matematica Italiana* (2), vol. 2 (1940), pp. 434-437.

[3] N. Steenrod, *Fibre Bundles*, Princeton, 1951.

[4] Cf., for instance, Chern, *Topics in Differential Geometry* (Mimeographed notes), Institute for Advanced Study, Princeton, 1951.

COMMUNTATORS OF OPERATORS.*

By Paul R. Halmos.

If H is a (complex) Hilbert space and if P and Q are operators on H (i. e. bounded linear transformations of H into itself), the *commutator* $[P, Q]$ of P and Q is defined by

$$[P, Q] = PQ - QP.$$

The *self-commutator* $[P]$ of a single operator P is defined by

$$[P] = [P^*, P] = P^*P - PP^*.$$

My purpose in this note is to make a slight contribution to our as yet very meager knowledge of what the commutator of two operators on a Hilbert space can look like. Wintner [3] proved that if P and Q are Hermitian, then $[P, Q]$ cannot be a non-zero multiple of the identity; as Putnam [1] has pointed out, Wintner's method yields the same conclusion even without the assumption that P and Q are Hermitian. Wielandt [2] obtained (by entirely different methods) a somewhat more general result, applicable to normed algebras. Wintner then asked whether or not the negative assertion that $[P, Q]$ can never be equal to the identity can be strengthened by proving that

$$\inf\{|([P, Q]x, x)| : \| x \| = 1\} = 0.$$

Putnam showed that this is always true on a finite dimensional Hilbert space and that it remains true in the infinite dimensional case if at least one of the two operators P and Q is Hermitian, or even normal, or even semi-normal. (An operator P is *semi-normal* if P^*P and PP^* are comparable with respect to the usual partial ordering of Hermitian operators.) I propose to show that, in general, the answer to Wintner's question is no. This assertion follows easily from the fact (Theorem 2) that the real (i. e., Hermitian) part of a commutator on an infinite dimensional Hilbert space may be prescribed arbitrarily. Theorem 2, in turn, is a consequence of the assertion (Theorem 1) that every Hermitian operator on an infinite dimensional Hilbert space is the sum of two self-commutators.

* Received May 25, 1951.

For a fixed Hilbert space H, let K be the set of all sequences $x = \{x_n\}$ such that $x_n \varepsilon H$, $n = 1, 2, \cdots$, and such that $\sum_n \| x_n \|^2 < \infty$. If, for any two elements x and y of K, the inner product of x and y is defined by

$$(x, y) = \sum_n (x_n, y_n),$$

then K is a Hilbert space; K is, in fact, the direct sum of countably many copies of H. Suppose that A is a Hermitian operator on H and define an operator B on K by $(Bx)_n = Ax_n$. Define another operator U on K by writing $(Ux)_1 = 0$ and $(Ux)_n = x_{n-1}$ for $n > 1$.

LEMMA 1. *If $P = BU$, then $([P]x)_1 = A^2x$ and $([P]x)_n = 0$ for $n > 1$.*

Proof. It is easy to verify that the operator B is Hermitian and that the adjoint of U is defined by $(U^*x)_n = x_{n+1}$. It follows that

$$(P^*Px)_n = (U^*B^2Ux)_n = (B^2Ux)_{n+1} = A^2(Ux)_{n+1} = A^2x_n$$

and, if $n > 1$, that

$$(PP^*x)_n = (BUU^*Bx)_n = A(UU^*Bx)_n = A(U^*Bx)_{n-1} = A(Bx)_n = A^2x_n.$$

Since $(PP^*x)_1 = A(UU^*Bx)_1 = A(0) = 0$, the proof of the lemma is complete.

It is convenient to say that a subspace H of a Hilbert space K is *large* if H contains infinitely many orthogonal copies of its orthogonal complement, or, in other words, if $\dim(H) \geqq \aleph_0 \dim(K - H)$. Thus, for example, a subspace of a separable Hilbert space is large if and only if it is infinite dimensional.

LEMMA 2. *A Hermitian operator with a large null space is a self-commutator.*

Proof. Suppose first that the given Hermitian operator is positive, i. e. that it can be written in the form A^2 with a Hermitian A. Let H be the closure of the range of A. Since H is the orthogonal complement of the null space of A, there is no loss of generality in assuming that the originally given Hilbert space $\tilde{H}$ contains the direct sum K of countably many copies of H, and that, moreover, H is embedded in K so that it coincides with the set of all those sequences x in K for which $x_n = 0$ whenever $n > 1$. If an operator P is defined on K, as in Lemma 1, and extended to $\tilde{H}$ by defining it to be 0 (or, for that matter, any normal operator) on the orthogonal complement $\tilde{H} - K$, then Lemma 1 implies the desired result. If the given operator is negative, the representation can be achievd with P^* in place of P. The case of a general Hermitian operator can be treated by putting together the results of the positive and the negative cases. It suffices to note that every Hermitian operator is the *direct* sum of a positive and a negative operator, and, in case the original operator has a large null space, then the direct summands can be selected so that they too have that property.

LEMMA 3. *Every Hermitian operator on an infinite dimensional Hilbert space leaves invariant at least one large subspace with a large orthogonal complement.*

Proof. The underlying Hilbert space, if it is not already separable, can be expressed as a direct sum of separable, infinite dimensional subspaces invariant under the given operator. There is, therefore, no loss of generality in restricting attention to separable Hilbert spaces. If A is Hermitian and E is the spectral measure of A, and if, for every Borel subset M of the real line, $E(M) = 0$ or 1, then A is a scalar multiple of 1. It follows easily that if, for every M, the dimension of the range of $E(M)$ is finite or co-finite, then A differs from a scalar multiple of 1 by a finite dimensional operator. In the contrary case both $E(M)$ and $1 - E(M)$ have infinite dimensional ranges for some M. In either case the conclusion of the lemma is obvious.

THEOREM 1. *Every Hermitian operator on an infinite dimensional Hilbert space is the sum of two self-commutators.*

Proof. By Lemma 3, the given operator is the sum of two Hermitian operators with large null spaces, and the theorem follows from Lemma 2.

To apply these results to a general operator P, it is necessary to break up P into its real and imaginary parts, i. e. the uniquely determined Hermitian operators A and B for which $P = A + iB$. If $P = A + iB$, $Q = C + iD$ (with A, B, C, and D Hermitian), it is convenient to write $P' = P'(P, Q) = A + iD$, $Q' = Q'(P, Q) = B + iC$. It follows that $P'' = P'(P', Q') = A + iC$, $Q'' = Q'(P', Q') = D + iB$, and finally that $P''' = P'(P'', Q') = P$, $Q''' = Q'(P'', Q'') = Q$. The reason for introducing P' and Q' is notational convenience; in terms of them it is easy to write down the commutator of P and Q. It is, in fact, a matter of automatic computation to verify that

$$[P, Q] = 2^{-1}([P'] + [Q']) + (2i)^{-1}([P''] + [Q'']).$$

Since a self-commutator is always Hermitian, and since an operator uniquely determines its real and imaginary parts, it follows that, for instance, the real part of $[P, Q]$ is $2^{-1}([P'] + [Q'])$. Since the transformation carrying P

and Q into P' and Q' is cyclic of order 3, it follows that any one of the three pairs $\{P, Q\}$, $\{P', Q'\}$, and $\{P'', Q''\}$ uniquely determines both others. These facts, combined with Theorem 1, yield the following result.

THEOREM 2. *Every Hermitian operator on an infinite dimensional Hilbert space is the real part of a commutator.*

COROLLARY. *There exist operators P and Q such that*

$$\inf\{|([P, Q]x, x)| : \| x \| = 1\} \geqq 1.$$

Proof. Theorem 2 yields the existence of two operators P and Q such that the real part of $[P, Q]$ is the identity. It follows that $([P, Q]x, x)$ where $\| x \| = 1$, is a complex number whose real part is 1, and that, consequently, $|([P, Q]x, x)| \geqq 1$.

It might be worth while, in closing, to call attention to another consequence of Theorem 1. Since a scalar multiple of a commutator is again a commutator, Theorem 1 and the decomposition of an operator into its real and imaginary parts imply that every operator on an infinite dimensional Hilbert space is the sum of four commutators. It follows that every additive functional of such operators, that vanishes on all commutators, vanishes identically, or, in other words, that the concept of *trace* cannot be extended to operators on infinite dimensional Hilbert spaces. (This comment was called to my attention by Irving Kaplansky.) Results of this type were known before, but only under additional assumptions of continuity or positiveness.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] C. R. Putnam, "On commutators of bounded matrices," this JOURNAL, vol. 73 (1951), pp. 127-131.

[2] H. Wielandt, "Ueber die Unbeschränktheit der Operatoren der Quantenmechanik," *Mathematische Annalen*, vol. 121 (1949), p. 21.

[3] A. Wintner, "The unboundedness of quantum-mechanical matrices," *Physical Review*, vol. 71 (1947), pp. 738-739.

ON HOMOTOPY GROUPS OF FUNCTION SPACES.*

By JAMES R. JACKSON.

1. Introduction. Let x_0 be a point of subset X_0 of topological space X, and let Y_0 be a subset of space Y. Let $y_0 \varepsilon Y_0$, and denote by the same symbol y_0 any constantly y_0-valued function. Throughout this paper, the following notations will be used:

Ω is the space of continuous mappings $f:(X, X_0) \to (Y, Y_0)$; that is, of mappings on X into Y which carry X_0 into Y_0.

Ω_0 is the space of continuous mappings $f:(X, X_0) \to (Y, y_0)$.

Ω_{00} is the space of continuous mappings $f:(X, X_0, x_0) \to (Y, Y_0, y_0)$.

Ψ is the space of continuous mappings $f: X_0 \to Y_0$.

Ψ_0 is the space of continuous mappings f: $(X_0, x_0) \to (Y_0, y_0)$.

We shall show (Section 10) that if X_0 is a retract of X, and if X, X_0, and Y satisfy certain rather general conditions (Sections 3, 4, and 5); then the m-th homotopy group $\Pi_m(\Omega, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi, y_0)$; and also that $\Pi_m(\Omega_{00}, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi_0, y_0)$.

(Group G is a *split extension of normal subgroup N by group H* if there exists a homomorphism of G onto H, with kernel N, and which induces an isomorphism of a subgroup H_0 of G onto H. It is well-known that if H_0 is also a normal subgroup, then G is the direct sum of H_0 and N.)

These results, together with some corollaries, enable us to relate the homotopy groups of many function-spaces to the homotopy groups of Y, and also to investigate certain homotopy classification problems. In particular, we provide a systematic approach to the structure of Fox's torus homotopy groups (Section 12), and list some miscellaneous interesting results (Section 13).

We also show (Section 8) that if X_0 is a deformation retract of X, and if X, X_0, and Y satisfy a weak restriction; then $\Pi_m(\Omega, y_0)$ is isomorphic to $\Pi_m(\Psi, y_0)$.

2. Some definitions. The subset $\{x \mid 0 \leq x_i \leq 1, i = 1, \cdots, m\}$ of

* Received June 6, 1951.

Euclidean m-space will be denoted by I^m. For I^1 we simply write I. The subset of point of I^m having at least one coordinate either zero or one will be designated by B^{m-1}.

Our definitions of homotopy, relative homotopy, homotopy groups, induced homomorphisms on homotopy groups, and other such concepts will be those of Fox [2] and Hu [4]. Our notation is essentially that of Hu.

Whenever a space of mappings is considered as a topological space, its topology will be the compact-open topology of Arens [1] and Fox [3].

3. Condition I. The condition discussed in the present section will be hypothesized in the main lemmas and theorems to follow. That it is not very restrictive is indicated by (3.4).

(3.1) *Definition.* We say that *X and Y satisfy Condition I* provided that whenever $\sigma: I^m \to Y^X$ is a continuous mapping, we may define a continuous mapping $\sigma^*: I^m \times X \to Y$ by

$$\sigma^*(t, x) = \sigma(t)(x) \qquad (t, x) \in I^m \times X. \tag{3.2}$$

Fox [3] has shown that for arbitrary X and Y, if $\sigma^*: I^m \times X \to Y$ is continuous, then (3.2) defines a continuous mapping $\sigma: I^m \to Y^X$. Thus, if X and Y satisfy Condition I, then (3.2) determines a one-one correspondence between the space of continuous functions on I^m into Y^X and the space of continuous functions on $I^m \times X$ into Y. Simple calculations show that this correspondence may be restricted to give a one-one correspondence between the continuous mappings $\sigma: (I^m, B^{m-1}) \to (\Omega, y_0)$ and the continuous mappings

$$\sigma^*: (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, Y_0, y_0).$$

One also concludes easily from Condition I that homotopies of the functions σ relative to $\{B^{m-1}, y_0\}$ are equivalent to homotopies of the corresponding functions σ^* relative to $\{I^m \times X_0, Y_0; B^{m-1} \times X, y_0\}$.

These considerations, with parallel ones concerning Ω_0, yield the following lemma, which has been informally stated and used by Hu[7] in a more restricted case.

(3.3) Lemma. *If X and Y satisfy Condition* I, *then the groups* $\Pi_m(\Omega, y_0)$ and $\Pi_m(\Omega_0, y_0)$, *respectively, may be considered to have as elements the homotopy classes relative to* $\{I^m \times X_0, Y_0; B^{m-1} \times X, y_0\}$ *of the continuous mappings*

$$\sigma^*: (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, Y_0, y_0),$$

and the homotopy classes relative to $\{I^m \times X_0, y_0; B^{m-1} \times X, y_0\}$ *of the continuous mappings*

$$\sigma^* : (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, y_0, y_0).$$

One sees easily that if α_1 and α_2 are each members of either of the groups mentioned in the lemma, and if they are represented by σ^*_1 and σ^*_2 according to the lemma, then the sum $\alpha_1 + \alpha_2$ is represented by $\sigma^*_3 = \sigma^*_3(t, x)$, defined by $\sigma^*_1(2t_1, t_2, \cdots, t_m, x)$ if $0 \leq t_1 \leq \frac{1}{2}$, and by $\sigma^*_2(2t_1 - 1, t_2, \cdots, t_m, x)$ if $\frac{1}{2} \leq t_1 \leq 1$.

The generality of Conditions I is indicated by the following theorem, a direct consequence of a theorem of Fox [3].

(3.4) THEOREM. *X and Y satisfy Condition* I *if either* (i) *X satisfies the first axiom of countability, or* (ii) *X is locally compact and regular* (*No restriction on Y in either case*).

4. Condition II. This is another hypothesis of our principal theorems. Its generality is indicated by (4.2).

Closed subset X_0 of space X is said to have the *homotopy extension property in X relative to space Y* if whenever continuous mappings $\phi : X \to Y$ and $\phi' : I \times X_0 \to Y$ satisfy $\phi'(0, x) = \phi(x)$ for $x \varepsilon X_0$; then ϕ' has a continuous extension $\phi'' : I \times X \to Y$ such that $\phi''(0, x) = \phi(x)$ for $x \varepsilon X$.

(4.1) *Definition.* We say that *X, X_0, and Y satisfy Condition II* provided that for $m = 1, 2, \cdots$, the subset $(I^m \times X_0) \cup (B^{m-1} \times X)$ of $I^m \times X$ has the homotopy extension property in $I^m \times X$ relative to Y.

The significance of this condition will appear in the proofs. The weakness of the restriction is indicated by the following theorem, which is a combination of [4, 9.2-9.5] with some standard theorems (For definitions of *ANR* and *ANR**, see [4]).

(4.2) THEOREM. *X, X_0, and Y satisfy Condition II if X_0 is closed in X, and if also any one of the four following requirements is met:*

(i) *X and X_0 ANR's.*

(ii) *X metric, Y an ANR.*

(iii) *X a Hausdorff space, $I^m \times X$ normal for $m = 1, 2, \cdots, Y$ an ANR*.*

(iv) *$I^m \times X$ normal for $m = 1, 2, \cdots, Y$ a compact ANR*.*

5. Conditions I and II. If X and Y, and also X_0 and Y satisfy Condition I, we say *X, X_0, and Y satisfy Condition I.* If also X, X_0. and Y satisfy Condition II, then we say that *X, X_0, and Y satisfy Conditions I and II.*

We shall mainly be interested in the case that X_0 is a retract of X (observe that this is always the case when X_0 reduces to a single point). To clear the air in this case, we set down a corollary which follows from (3.4), (4.2), and some standard theorems.

(5.1) THEOREM. *Let X_0 be a retract of X. Then X, X_0, and Y satisfy Conditions I and II if either* (i) *X is an ANR, or* (ii) *X is metric and Y is an ANR.*

These conditions are of special interest since every locally-finite polyhedron is an *ANR* [8].

6. Some homomorphisms. Define $\theta:\Omega\rightarrow\Psi$ by $\theta(\phi)=\phi|X_0$, $\phi\,\varepsilon\,\Omega$. Define $j:\Omega_0\rightarrow\Omega$ by $j(\phi)=\phi$, $\phi\,\varepsilon\,\Omega_0$. The functions θ and j are obviously continuous.

(6.1) LEMMA. *If X, X_0, and Y satisfy Conditions I and II, then the induced homomorphism $j^*:\Pi_m(\Omega_0,y_0)\rightarrow\Pi_m(\Omega,y_0)$ carries $\Pi_m(\Omega_0,y_0)$ onto the kernel of the induced homomorphism $\theta^*:\Pi_m(\Omega,y_0)\rightarrow\Pi_m(\Psi,y_0)$.*

Proof of (6.1). It is obvious that the image of j^* is contained in the kernel of θ^*. Hence we must show that if $\alpha\,\varepsilon\,\Pi_m(\Omega,y_0)$, and if $\theta^*(\alpha)=0$, then for some $\beta\,\varepsilon\,\Pi_m(\Omega_0,y_0)$, we have $\alpha=j^*(\beta)$. Lemma (3.3) reduces this to the following proposition.

(6.2) Let $\sigma:(I^m\times X, I^m\times X_0, B^{m-1}\times X)\rightarrow(Y,Y_0,y_0)$ be continuous, and suppose there exists a mapping $\sigma':I\times I^m\times X_0\rightarrow Y_0$ such that $\sigma'(0,t,x)=\sigma(t,x)$, $(t,x)\,\varepsilon\,I^m\times X_0$; $\sigma'(I\times B^{m-1}\times X_0)=y_0=\sigma'(1\times I^m\times X_0)$. Then there exists a mapping $\sigma'':I\times I^m\times X\rightarrow Y$ such that

$$\sigma''(0,t,x)=\sigma(t,x),\ (t,x)\,\varepsilon\,I^m\times X;\qquad \sigma''(I\times I^m\times X_0)\subset Y_0$$
$$\sigma''(I\times B^{m-1}\times X)=y_0=\sigma''(1\times I^m\times X_0).$$

For suppose (6.2) is true. If σ represents an element α of the kernel of θ^*, then the hypotheses of (6.2) are fulfilled, whence σ'' exists. Define $\sigma_1:I^m\times X\rightarrow Y$ by $\sigma_1(t,x)=\sigma''(1,t,x)$, $(t,x)\,\varepsilon\,I^m\times X$. Clearly σ_1 represents α. Since $\sigma_1(I^m\times X_0)=y_0$, the function σ_1 also represents some element $\beta\,\varepsilon\,\Pi_m(\Omega_0,y_0)$. One sees easily from the definition of j^* that $j^*(\beta)=\alpha$, as required.

Proof of (6.2). Extend σ' to $I \times [(I^m \times X_0) \cup (B^{m-1} \times X)]$ by setting $\sigma'(I \times B^{m-1} \times X) = y_0$. Obviously the extended σ' is continuous and satisfies

$$\sigma'(0, t, x) = \sigma(t, x), \qquad (t, x) \in (I^m \times X_0) \cup (B^{m-1} \times X).$$

Hence by Condition II, σ' has an extension $\sigma'' : I \times I^m \times X \to Y$ such that $\sigma''(0, t, x) = \sigma(t, x)$, $(t, x) \in I^m \times X$. The function σ'' obviously satisfies the requirements of the conclusion of (6.2).

7. An important lemma. In this section we do not need Condition I or Condition II, but we shall require the more restrictive hypothesis that X_0 be a retract of X. Lemma (7.1)—in a restricted form suggested by the fact that Y may be considered as a retract of Y^X—was the starting point of the present investigation.

Let $\rho_0 : X \to X_0$ be a retraction of X onto X_0. Define $\rho : \Psi \to \Omega$ by $\rho(\phi) = \phi\rho_0$, $\phi \in \Psi$. One sees easily that ρ is continuous. Let $\rho^* : \Pi_m(\Psi, y_0) \to \Pi_m(\Omega, y_0)$ be the homomorphism induced by ρ.

(7.1) LEMMA.

(i) $\theta^*\rho^* : \Pi_m(\Psi, y_0) \to \Pi_m(\Psi, y_0)$ *is the identity automorphism.*

(ii) $\rho^* : \Pi_m(\Psi, y_0) \to \Pi_m(\Omega, y_0)$ *is an isomorphism into.*

(iii) $\theta^* : \Pi_m(\Omega, y_0) \to \Pi_m(\Psi, y_0)$ *is an onto homomorphism.*

Proof of (7.1). Conclusion (i) is the fruit of a simple calculation; (ii) and (iii) are elementary set-theoretical consequences of (i).

8. An isomorphism theorem. The following result, which is not needed for the ensuing theory, has obvious generalizations relating to the concept of homotopy type.

(8.1) THEOREM. *If X, X_0, and Y satisfy Condition I, and if X_0 is a deformation retract of X; then $\theta^* : \Pi_m(\Omega, y_0) \to \Pi_m(\Psi, y_0)$ is an isomorphism onto.*

Proof of (8.1). By (7.1)(iii), we need only show that θ^* has kernel 0. Let $\tau : I \times X \to X$ retract X onto X_0 by deformation. Lemma (3.3) reduces (8.1) to the following proposition.

(8.2) Let $\sigma : (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, Y_0, y_0)$, and suppose there exists a mapping $\sigma' : I \times I^m \times X_0 \to Y_0$ such that

$$\sigma'(0, t, x) = \sigma(t, x), \qquad (t, x) \in I^m \times X_0;$$
$$\sigma'(I \times B^{m-1} \times X_0) = y_0 = \sigma'(1 \times I^m \times X_0).$$

Then there exists a mapping $\sigma'': I \times I^m \times X \to Y$ such that

$$\sigma''(0, t, x) = \sigma(t, x), \ (t, x) \in I^m \times X; \qquad \sigma''(I \times I^m \times X_0) \subset Y_0;$$
$$\sigma''(I \times B^{m-1} \times X) = y_0 = \sigma''(1 \times I^m \times X).$$

Proof of (8.2). Define $\sigma''(s, t, x)$ to be $\sigma(t, \tau(2s, x))$ if $0 \leq s \leq \frac{1}{2}$ and $\sigma'(2s-1, t, \tau(1, x))$ if $\frac{1}{2} \leq s \leq 1$. It is easily verified that σ'' satisfies the requirements of the conclusion of (8.2).

We set down ahead of time a simple corollary of (8.1) and (9.1).

(8.3) COROLLARY. *Let X, X_0, and Y satisfy Conditions I and II, and suppose X_0 is a deformation retract of X. Then $\Pi_m(\Omega_0, y_0)$ is trivial (that is, consists of a single element).*

9. The main lemma. The following result sharpens (6.1) for the case that X_0 is a retract of X.

(9.1) LEMMA. *If X, X_0, and Y satisfy Conditions I and II, and if X_0 is a retract of X; then the homomorphism j^* carries $\Pi_m(\Omega_0, y_0)$ isomorphically onto the kernel of θ^*.*

Proof of (9.1). By (6.1), we need only show that j^* has kernel 0. Let $\rho: X \to X_0$ be a retraction of X onto X_0. By means of (3.3), we reduce (9.1) to the following proposition.

(9.2) Let $\sigma: (I^m \times X, I^m \times X_0, B^{m-1} \times X) \to (Y, y_0, y_0)$ be continuous, and suppose there exists a mapping $\sigma': I \times I^m \times X \to Y$ such that

$$\sigma'(0, t, x) = \sigma(t, x), \ (t, x) \in I^m \times X; \qquad \sigma'(I \times I^m \times X_0) \subset Y_0;$$
$$\sigma'(I \times B^{m-1} \times X) = y_0 = \sigma'(1 \times I^m \times X).$$

Then there exists a mapping $\sigma'': I \times I^m \times X \to$ Y satisfying the same requirements, and also $\sigma''(I \times I^m \times X_0) = y_0$.

Proof of (9.2). Define

$$\begin{aligned}
&\tau(r, s, t, x) = y_0, && r \in I, 0 \leq s \leq r, t \in I^m, x \in X_0;\\
&\tau(r, s, t, x) = \sigma'(s-r, t, x), && r \in I, r \leq s \leq 1, t \in I^m, x \in X_0;\\
&\tau(I \times I \times B^{m-1} \times X) = y_0;\\
&\tau(r, 0, t, x) = \sigma'(0, t, x), && r \in I, t \in I^m, x \in X;\\
&\tau(r, 1, t, x) = \sigma'(1-r, t, \rho(x)), && r \in I, t \in I^m, x \in X.
\end{aligned}$$

It is easily seen that τ is single-valued and continuous, and is defined on $I \times [(I^{m+1} \times X_0) \cup (B^m \times X)]$. Hence by Condition II, it has an extension $\tau': I \times I^{m+1} \times X = I \times I \times I^m \times X \to Y$. Define $\sigma''(s, t, x) = \tau'(1, s, t, x)$, $(s, t, x) \in I \times I^m \times X$. One sees without difficulty that σ'' satisfies the requirements of the conclusion of (9. 2).

10. The main theorems. The first theorem of this section is merely a combination of (7. 1) and (9. 1). The second theorem is a purely algebraic consequence of the first.

(10. 1) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then $\Pi_m(\Omega, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi, y_0)$.*

(10. 2) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then:*

(i) *$\Pi_m(\Omega, y_0)$ has a subgroup isomorphic to $\Pi_m(\Psi, y_0)$, and a normal subgroup isomorphic to $\Pi_m(\Omega_0, y_0)$.*

(ii) *If either of $\Pi_m(\Psi, y_0)$ and $\Pi_m(\Omega_0, y_0)$ consists of a single element, then $\Pi_m(\Omega, y_0)$ is isomorphic to the other.*

(iii) *If $\Pi_m(\Omega, y_0)$ is abelian, then it is isomorphic to the direct sum $\Pi_m(\Psi, y_0) + \Pi_m(\Omega_0, y_0)$.*

Note that the hypothesis of (10. 2) (iii) is always fulfilled for $m \geqq 2$, and is fulfilled for all $m \geqq 1$ if Y is a topological group and Y_0 a subgroup [6].

Following through the proofs on which (10. 1) is based, one sees easily that the following theorem can be established in exactly the same way.

(10. 3) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then $\Pi_m(\Omega_{00}, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega_0, y_0)$ by $\Pi_m(\Psi_0, y_0)$.*

Theorem (10. 3) has, of course, a corollary parallel to (10. 2).

We now provide a theorem concerning the relative homotopy groups $\Pi_m(\Omega, \Omega_0, y_0)$ and $\Pi_m(\Omega_{00}, \Omega_0, y_0)$, where Ω_0 is considered as a subset of Ω and of Ω_{00} in the obvious way. The proof, which follows easily from (7. 1) (iii), (9. 1), and the exactness of the homotopy sequence, is omitted. For relevant definitions and theorems, see [2] or [4, pp. 80-82, 94-99].

(10. 4) THEOREM. *Let X, X_0, and Y satisfy Conditions I and II, and let X_0 be a retract of X. Then for $m = 2, 3, \cdots$, $\Pi_m(\Omega, \Omega_0, y_0)$ is isomorphic to $\Pi_m(\Psi, y_0)$, and $\Pi_m(\Omega_{00}, \Omega_0, y_0)$ is isomorphic to $\Pi_m(\Psi_0, y_0)$.*

Those familiar with relative homotopy theory and the homotopy sequence might suspect that (10. 1) and (10. 3) might be more easily established by first proving (10. 4) directly. However, this does not seem to be the case, since existing theorems do not apply conveniently to the lemmas which arise in setting up a direct proof of (10. 4).

11. Factor spaces. The theorems of Section 10 do not lend themselves to calculations of any generality because of the appearance of the group $\Pi_m(\Omega_0, y_0)$, which is difficult to deal with directly. We get around this by introducing the *factor space* $X^* = X/X_0$, which is the space obtained from X by identifying the points of X_0 to the single point x^*. The map of identification $m_0 : X \to X^*$ is defined by $m_0(x) = x$, $x \varepsilon X - X_0$, and $m_0(X_0) = x^*$. The set X^* is topologized by taking subset U open if and only if $m_0^{-1}(U)$ is an open subset of X [9].

Let Ω^*_0 be the space of continuous functions $f : (X^*, x^*) \to (Y, y_0)$. It is not difficult to see that if X_0 is compact, then a homeomorphism $m : \Omega^*_0 \to \Omega_0$ is defined by setting $m(\phi) = \phi m_0$, $\phi \varepsilon \Omega^*_0$. Then the following theorem is obvious.

(11. 1) THEOREM. *Let X_0 be compact. Then $\Pi_m(\Omega_0, y_0)$ is isomorphic to $\Pi_m(\Omega^*_0, y_0)$.*

We shall not enumerate the obvious corollaries to the theorems of Section 10.

12. An application. In this section we shall apply (10. 1) and (11. 1) to obtain the results of Fox [2] concerning the algebraic structure of his torus homotopy groups.

(12. 1) DEFINITION. *Let T^0 be a point, and for $r = 1, 2, \cdots$, let T^r be the r-fold topological product of 1-spheres. Define $\tau_r{}^m(Y, y_0) = \Pi_m(Y^{T^{r-1}}, y_0)$, $m, r = 1, 2, \cdots$.*

Fox pointed out that $\tau_r{}^1(Y, y_0)$ *is identical with his r-th torus homotopy group of Y at base-point y_0.*

Consider T^{r-1} to be parametrized by $r - 1$ real numbers modulo 1. Define $\rho : T^{r-1} \to T^{r-1}$, $(r \geq 2)$, by $\rho(x) = (0, x_2, \cdots, x_{r-1})$, $x \varepsilon T^{r-1}$. Plainly ρ is a retraction of T^{r-1} onto a (compact) subset homeomorphic to T^{r-2}. Hence by (10. 1) and (11. 1), we conclude that $\tau_r{}^m(Y, y_0)$ is isomorphic to a split extension of $\Pi_m(\Omega^*_0, y_0)$ by $\tau^m{}_{r-1}(Y, y_0)$, where we see without difficulty that $\Pi_m(\Omega^*_0, y_0)$ is isomorphic to $\Pi_{m+1}(Y^{T^{r-2}}, y_0)$. This proves the following proposition.

(12.2) *$\tau_r^m(Y, y_0)$ is isomorphic to a split extension of $\tau_{r-1}^{m+1}(Y, y_0)$ by $\tau_{r-1}^m(Y, y_0)$, for $r \geq 2$.*

Now $\tau_{r-1}^{m+1}(Y, y_0)$ is commutative, so we can apply (12.2) to it to obtain a direct sum decomposition. Remembering that T^0 is a single point, so that $\tau_1^m(Y, y_0)$ is isomorphic to $\Pi_m(Y, y_0)$, we obtain the following direct sum decomposition by a simple induction.

$$(12.3) \qquad \tau_{r-1}^{m+1}(Y, y_0) \simeq \sum_{j=2}^{r} \tbinom{r-2}{j-2} \Pi_{m+j-1}(Y, y_0).$$

Fox's structure theorem is obtained by combining (12.2) and (12.3) for the case $m = 1$.

13. Further applications. Many of the known results on homotopy groups of spaces of inessential functions are easy corollaries to the theorems of Sections 10 and 11, which also open to investigation the homotopy theories of many functions spaces inaccessible to existing results.

In view of (3.3), it is clear that any theorem on homotopy groups of function spaces can be interpreted as a homotopy classification theorem of a special type.

We give some simple applications of our theorems. The first two are slight generalizations of known results [5], but are included for completeness, since they are needed for the other applications.

Let S^k be the k-sphere ($k \geq 0$: S^0 is a pair of points), and let s_k be a fixed point of S^k.

$$(13.1) \qquad \Pi_m(Y^{S^k}\{s_k, y_0\}, y_0) \simeq \Pi_{m+k}(Y, y_0).$$

(13.2) $\Pi_m(Y^{S^k}\{s_k, Y_0\}, y_0)$ *is isomorphic to a split extension of* $\Pi_{m+k}(Y, y_0)$ *by* $\Pi_m(Y_0, y_0)$.

Proof of (13.1). By (11.1), we have, for $k > 0$,

$$\Pi_m(Y^{S^k}\{s_k, y_0\}, y_0) \simeq \Pi_m(Y^{I^k}\{B^{k-1}, y_0\}, y_0).$$

That

$$\Pi_m(Y^{I^k}\{B^{k-1}, y_0\}, y_0) \simeq \Pi_{m+k}(Y, y_0),$$

is an immediate consequence of (3.3). The case $k = 0$ is trivial.

Proof of (13.2). Take $X = S^k$ and $X_0 = s_k$ in (10.1). (13.2) follows at once from (13. 1) and the obvious fact that $Y_0{}^{s_k}$ is essentially identical with Y_0.

The following result has numerous obvious generalizations, whose proofs are similar to that below.

(13. 3) *Let X be the union of S^{i-1} and S^{j-1}, joined together by identifying the points s_{i-1} and s_{j-1} to a single point x_0. Then*

$$\Pi_1(Y^X\{x_0, y_0\}. y_0) \simeq \Pi_i(Y, y_0) + \Pi_j(Y, y_0).$$

Proof of (13. 3). Applying (10. 3) with $X_0 = S^{i-1}$, (11. 1), and (13. 1), we see that $\Pi_1(Y^X\{x_0, y_0\}, y_0)$ is isomorphic to a split extension of $\Pi_i(Y, y_0)$ by $\Pi_j(Y, y_0)$. Another application of the some theorems with $X_0 = S^{j-1}$ shows that the isomorphic image of $\Pi_j(Y, y_0)$ in $\Pi_1(Y^X\{x_0, y_0\}, y_0)$ is a normal subgroup, whence the proposition follows.

We state one simple result concerning homotopy classification. Its proof follows from (13. 2), when we observe that the function space with which it is concerned is homeomorphic to the space of representatives given by (3. 3) for $\Pi_1(Y^{S^k}\{s_k, Y_0\}, y_0)$.

(13. 4) *The homotopy classes relative to $\{s_k \times S^1, Y_0; S^k \times s_1, y_0\}$ of the space of continuous functions*

$$f:(S^k \times S^1, s_k \times S^1, S^k \times s_1) \to (Y, Y_0, y_0)$$

can be put in a 1-1 correspondence with some split extension of $\Pi_{k+1}(Y, y_0)$ by $\Pi_1(Y_0, y_0)$, and hence with the direct sum $\Pi_{k+1}(Y, y_0) + \Pi_1(Y_0, y_0)$.

The preceding applications of this section are almost obvious intuitively, although a rigorous proof is in each case except (13. 1) rather difficult, without the theorems of Sections 10 and 11. The following proposition is less accessible to the imagination.

(13. 5) *Let X_p be the closed orientable surface of genus p. Then $\Pi_m(Y^{X_p}, y_0) = G_0$ has a normal series of subgroups $G_0 \supset G_1 \supset \cdots \supset G_{2p+1}$, where:*

(i) *G_0 is isomorphic to a split extension of G_1 by $\Pi_m(Y, y_0)$;*

(ii) *G_i is isomorphic to a split extension of G_{i+1} by $\Pi_{m+1}(Y, y_0)$, for $i = 1, \cdots, 2p$;*

(iii) $G_{2p+1} \simeq \Pi_{m+2}(Y, y_0)$.

Proof of (13.5). The case $p=0$ is contained in (13.2). If $p>0$, consider X_p as a sphere with p handles. Single out a handle H, and parametrize it in the obvious way by longitude θ $(-\pi/2 \leq \theta \leq \pi/2)$ and latitude ϕ $(-\pi < \phi \leq \pi)$. Designate the point whose coordinates are θ and ϕ by (θ, ϕ).

In (10.1), take X_0 the single point $(0,0) \varepsilon H$. We find that G_0 is isomorphic to a split extension of

$$(13.6) \qquad \Pi_m(Y^{X_p}\{(0,0), y_0\}, y_0) = G'_1$$

by a group essentially identical with $\Pi_m(Y, y_0)$.

We may construct X_p in such a way that $(-\pi/2, 0)$ can be joined in X_p to $(\pi/2, 0)$ by an arc of a great circle in the surface of the sphere. Let X_0 be the union of this arc with the points of H for which $\phi = 0$. Then $(0,0) \varepsilon X_0 \subset X_p$, and examination of the figure makes it clear that X_0 is a 1-sphere and is a retract of X_p. Then by (10.3) and (13.1), G'_1 is isomorphic to a split extension of $\Pi_m(Y^{X_p}\{X_0, y_0\}, y_0) = G'_2$ by $\Pi_{m+1}(Y, y_0)$.

Let X^*_p be the space obtained from X_p by identifying the points of X_0 to the single point x^*. By (11.1), G'_2 is isomorphic to $\Pi_m(Y^{X^*_p}\{x^*, y_0\}, y_0)$. Let X^*_0 be the image in X^*_p under the map of identification of the set of points in H such that $\theta = 0$. One sees easily from a sketch that X^*_0 is a 1-sphere and is a retract of X^*_p. By (10.3) and (13.1), G'_2 is isomorphic to a split extension of $\Pi_m(Y^{X^*_p}\{X^*_0, y_0\}, y_0) = G'_3$ by $\Pi_{m+1}(Y, y_0)$. Using (11.1) we see easily that if $x \varepsilon X_{p-1}$, then G'_3 is isomorphic to $\Pi_m(Y^{X_{p-1}}\{x, y_0\}, y_0)$.

Thus. in essence, we are back to (13.6), except that p is reduced by unity. The proof is completed by a simple induction, concluding in an appeal to the case $p=0$, and followed by the identification of the groups G'_i with subgroups of G_0.

We note in closing that the theorems of Section 10, by reducing problems concerning spaces like Ω and Ω_{00} to problems concerning Ψ, Ψ_0, and Ω_0, greatly extend the range of application of the theorems recently obtained by Hu [7], relating homotopy groups of certain function spaces to certain subgroups of cohomology groups. Our principal theorems are, incidentally, generalizations of certain theorems of [7], but were deduced without knowledge of that paper.

UNIVERSITY OF CALIFORNIA,
LOS ANGELES.

BIBLIOGRAPHY.

[1] Arens, R. F., "A topology for spaces of transformations," *Annals of Mathematics*, vol. 47 (1946), pp. 480-495.

[2] Fox, R. H., "Homotopy groups and torus homotopy groups," *ibid.*, vol. 49 (1947), pp. 471-510.

[3] ———, "On topologies for function spaces," Bulletin of the American Mathematical Society, vol. 51 (1945), pp. 429-432.

[4] Hu, S. T., "Homotopy Theory," vol. 1, Dittoed typescript prepared by the Department of Mathematics, Tulane University, under Office of Naval Research Contract N7-ONR-434 Task Order III, 1950.

[5] ———, "On spherical mappings in a metric space," *Annals of Mathematics*, vol. 48 (1947), pp. 717-734.

[6] ———, "Some homotopy properties of topological groups and homogeneous spaces," *ibid.*, vol. 49 (1948), *pp.* 67-74.

[7] ———, "Structure of the homotopy groups of mapping spaces," *American Journal of Mathematics*, vol. 71 (1949), pp. 574-586.

[8] Liao, S. D., "On non-compact absolute neighborhood retracts," *Academia Sinica Science Record*, vol. 2 (1949), pp. 249-262.

[9] Seifert, H., and Threlfall, W., *Lehrbuch der Topologie*, Leipzig, 1934 (especially pp. 31-35).

A REMARK ON ISOLATED CRITICAL POINTS.*

By Erich H. Rothe.

1. Introduction. Let $I = I(x)$ be a real valued function of the point x of a space E to be specified later. We assume that the zero point o of E is an isolated critical point, i. e., that

$$(1.1) \qquad \operatorname{grad} I(x) = o \qquad \text{for } x = o$$

while grad $I \neq o$ for all $x \neq o$ of some neighborhood of o. In many investigations about critical points the following property which we formulate as a "hypothesis" plays a decisive role:

Hypothesis H. There exists a neighborhood U of o such that for all $x \neq o$ of the intersection [1] $U \wedge \{I(x) = I(o)\}$ the vectors $x - o$ and grad $I(x)$ are linearly independent.

If E is the (real) Euclidean n-space E^n, H is known to be true under either of the following two conditions: (i) o is a non-degenerate critical point [2]; (ii) $I(x)$ is analytic in the neighborhood of o.[3]

In a recent paper [4] hypothesis H serves as the main assumption of the theorem that for $E = E^n$ the alternating sum of the type numbers of the critical point equals the index of the singularity o of the vector field grad $I(x)$, and without proof it has been stated in this paper that the following condition is sufficient for the validity of the hypothesis H: there is an integer $p \geqq 2$ such that I has continuous differentials up to and including order $p + 2$; all differentials of order less than p vanish at $x = o$, while the homogeneous form of degree p giving the p-th differential at $x = o$ is not degenerate in the algebraic sense. This condition will be called "non-degeneracy of order p" [5] since for $p = 2$ it coincides essentially with the customary non-degeneracy condition.

* Received May 16, 1950.

[1] The symbol $\wedge$ denotes intersection, and for any property P, the symbol $\{P(x)\}$ denotes the set of all x having the property P.

[2] [6], p. 155, Theorem 4. 2.

[3] [1], Lemma 10; [6], p. 156, Theorem 4. 3.

[4] [7].

[5] Definition 2. 6.

The object of the present paper is then to prove the theorem that non-degeneracy of order p is sufficient for the validity of the hypothesis H (Theorem 3.1). The proof will be given for the case that E is a (not necessarily separable) Hilbert space which of course includes the case of a Euclidean *n-space*. The proof is given in Section 3 while Section 2 contains preliminary definitions and facts about differentials and gradients in a Hilbert space not all of which are new.

2. On differentials and gradients in the Hilbert space *E*. Let E be a Hilbert space, i. e., a real Banach space in which for any couple x, y of elements, a scalar product (x, y) is defined which satisfies the usual rules and is such that $(x, x)^{\frac{1}{2}}$ is the norm $\| x \|$ of the element x of E.

Definition 2.1. Let $f(x)$ be a continuous map of some open convex subset C of E into a Hilbert space E_1.[6] We define inductively differentials $d^0, d^1, d^2, \cdots$ of f as follows: $d^0 f(x) = f(x)$ for $x \varepsilon C$. Suppose now that for some integer $i \geqq 1, d^{i-1} = d^{i-1}f(x, h_1, h_2, \cdots, h_{i-1})$ has been defined for all i-tuples $(x, h_1, h_2, \cdots, h_{i-1})$ of elements of E such that $x + h_1 + h_2 + \cdots + h_{i-1} \varepsilon C$. The i-th differential d^i is then defined if and only if there exists a map $d^i = d^i f(x, h_1, \cdots, h_{i-1}, h_i)$ mapping all those $(i+1)$-tuples $(x, h_1, h_2, \cdots, h_{i-1}, h_i)$ of elements of E such that $x + h_1 + \cdots + h_{i-1} + h_i \varepsilon C$ into E_1 which has the following properties: d^i is linear[7] in h_i and

$$(2.1) \quad d^{i-1}f(x + h_i, h_1, \cdots, h_{i-1}) - d^{i-1}f(x, h_1, \cdots, h_{i-1}) = d^i f(x, h_1, \cdots, h_{i-1}, h_i) + \epsilon(x, h_1, \cdots, h_{i-1}, h_i)$$

with

$$(2.2) \quad \lim \epsilon(x, h_1, \cdots, h_{i-1}, h_i)/\| h_i \| = 0, \text{ where } \| h_i \| \to 0.$$

If a $d^i f(x, h_1, \cdots, h_{i-1}, h_i)$ with these properties exists it is uniquely determined[8] and is called the i-th differential d^i of f in C. $d^i f(x, h_1, \cdots, h_i)$ is i-linear in $h_1, \cdots, h_i$. In addition we have the important

LEMMA 2.1. *If $d^i f(x, h_1, h_2, \cdots, h_i)$ is continuous in the argument x for all x of a neighborhood of the point x_0 then it is symmetric in $h_1, h_2, \cdots, h_i$ at $x = x_0$.*

[6] In this paper $f(x)$ will be either a real valued function (i. e., E_1 the real line) or a map of C into $E_1 = E$.

[7] "Linear" means additive *and continuous*.

[8] [3], Lemma 11.1.

For the proof see [2], Theorem 8 or [4], Satz 1.

Definition 2.2. If $Q(h_1, h_2, \cdots, h_i)$ is an i-linear symmetric real function we set $Q(h) = Q(h_1, h_2, \cdots, h_i)_{h_1=h_2\ldots=h_i=h}$, and call $Q(h)$ a form of degree i. Correspondingly, if $f(x)$ is real-valued (i. e., E_1 the real axis) and $d^if(x, h_1, \cdots, h_i)$ exists in C we define $d^if(x, h)$ for $x + ih \varepsilon C$, by setting $d^if(x, h) = d^if(x, h_1, h_2, \cdots, h_i)_{h_1=h_2\ldots=h_i=h}$.

If f has continuous differentials up to and including order $(n+1)$ in C then the Taylor formula

$$(2.3) \qquad f(x+h) - f(x) = \sum_{j=1}^{n} d^jf(x, h)/j! + R_{n+1};$$

$$(2.4) \qquad R_{n+1} = \int_0^1 (1-t)^n/n!\,d^{n+1}f(x+th, h)\,dt$$

holds if x and $x + h$ are in C.[9]

So far we have used only the Banach space property of E. We now make use of the basic property of a Hilbert space E that to every linear functional $l(x)$ there exists a uniquely determined element $g \varepsilon E$ such that $l(x) = (g, x)$ where (g, x) denotes the scalar product of the elements g and x. This property together with the linearity of d^if in h_i makes the following definition possible:

Definition 2.3. If the real-valued $f(x)$ has an i-th differential $d^if(x, h_1, \cdots, h_i)$ then the i-th gradient $g^i = g^i(x, h_1, \cdots, h_{i-1})$ is defined as the element of E which is uniquely determined by the equation

$$(2.5) \qquad d^if(x, h_1, \cdots, h_{i-1}, h_i) = (g^i(x, h_1, \cdots, h_{i-1}), h_i) = (g^i, h_i).$$

The i-linearity of d^if implies obviously the $(i-1)$-linearity of $g^i(x, h_1, \cdots, h_{i-1})$ in $h_1, \cdots, h_{i-1}$. In particular, the first gradient function $g^1(x)$ is called the gradient of $f(x)$ and we write

$$(2.6) \qquad g^1(x) = g(x) = \operatorname{grad} f(x);$$

$$(2.7) \qquad g^i(x, h_1, \cdots, h_{i-1}) = g^i(x, h) \text{ if } h_1 = \cdots = h_{i-1} = h.$$

LEMMA 2.2. *If $d^if(x, h)$ exists and is continuous in x at $x = x_0$, then there exists a neighborhood U of x_0 and a constant $C = C(x_0)$ such that for all $x \varepsilon U$*

---

[9] [2], Theorem 5. The integrals are "Riemann" integrals in the sense defined in [2], p. 166.

$$\| d^i f(x, h_1, \cdots, h_i) \| \leqq C \| h_1 \| \cdots \| h_{i-1} \| \| h_i \|,$$

$$\| g^i (x, h_1, \cdots, h_{i-1}) \| \leqq C \| h_1 \| \cdots \| h_{i-1} \|.$$

Proof. The first inequality follows from [4], Hilfssatz 2. The second inequality follows from the first by setting $h_i = g^i(x, h_1, \cdots, h_{i-1})$ in (2. 5).

LEMMA 2. 3. (a) *The i-th gradient $g^i(x, h_1, \cdots, h_{i-1})$ (Definition 2. 3) is symmetric in $h_1, \cdots, h_{i-1}$ if g^i is continuous in x.* (b) $d^{i-1}g(x, h) = g^i(x, h)$.

Proof. Since every permutation of $h_1, \cdots, h_{i-1}$ may be considered as a permutation of $h_1, \cdots, h_{i-1}, h_i$ (leaving h_i fixed), Lemma 2. 3(a) is an immediate consequence of (2. 5) and Lemma 2. 1.

To prove (b) we note that, because of the symmetry in $h_1, \cdots, h_{i-1}$,

$$d^i f(x, h_1, \cdots, h_{i-1}, h_i) = (g^i(x, h_1, \cdots, h_{i-2}, h_i), h_{i-1}),$$

$$d^{i-1} f(x, h_1, \cdots, h_{i-1}) = (g^{i-1}(x, h_1, \cdots, h_{i-2}), h_{i-1}).$$

Because of $d^i = dd^{i-1}$ it follows easily from the second equation that

$$d^i f(x, h_1, \cdots, h_{i-1}, h_i) = (dg^{i-1}(x, h_1, \cdots, h_{i-2}, h_i), h_{i-1})$$

and comparison with the first equation proves

$$g^i(x, h_1, \cdots, h_{i-2}, h_i) = dg^{i-1}(x, h_1, \cdots, h_{i-2}, h_i).$$

Recursive application of this formula yields (b).

Definition 2. 4. If $P(h_1, \cdots, h_{i-1})$ is an element of E which is symmetric and linear in the h_j, we set $P(h) = P(h_1, \cdots, h_{i-1})_{h_1 = \cdots = h_{i-1} = h}$ and call $P(h)$ a polynomial of degree $i - 1$ in h.[10]

For later reference we write the Taylor formula (2. 3), (2. 4) in terms of gradients (Definition 2. 3)

$$f(x + h) - f(x) = \sum_{j=1}^{n} (g^j(x, h), h)/j! + R_{n+1}(x, h); \tag{2. 8}$$

$$R_{n+1}(x, h) = \int_0^1 (1 - t)^n / n! (g^{(n+1)}(x + th, h), h) dt. \tag{2. 9}$$

It is easily seen that a form $Q(h)$ of degree i (Definition 2. 2) has a gradient. We define:

---

[10] [5], p. 63. The original definition goes back to Banach.

Definition 2.5. The form $Q(h)$ is non-degenerate if grad $Q(h) \neq o$ for $h \neq o$. If there exists a constant $m > 0$ such that

$$\| \operatorname{grad} Q(h) \| \geqq m \qquad \text{for } \| h \| = 1, \tag{2.10}$$

$Q(h)$ is called strictly non-degenerate.

Remark to Definition 2.5. If E is the Euclidean n-space E^n the above definition of non-degeneracy coincides with the usual one: let $h^1, h^2, \cdots, h^n$ be the components of h in some coordinate system; then $Q(h)$ is degenerate if and only if the equations $\partial Q/\partial h_i = 0$ $(i = 1, 2, \cdots, n)$ have $h^1 = h^2 = \cdots = h^n = 0$ as the only common solution. In this case a non-degenerate form is obviously also strictly non-degenerate.

The following Lemmas 2.4 and 2.5 state some simple properties of differentials and gradients. We omit their simple proofs (cf. [3], p. 138).

LEMMA 2.4. *Let $f(h_1, h_2, \cdots, h_i)$ be a function of the i elements $h_1, \cdots, h_i$ of E. We assume that for $j = 1, 2, \cdots, i$ the differentials $d_jf(h_1, h_2, \cdots, h_i)$ of f with respect to h_j exist and are continuous in $(h_1, \cdots, h_i)$. Moreover, let $g_j(h_1, \cdots, h_i) = \operatorname{grad}_j f(h_1, \cdots, h_i)$ denote the gradient of $f(h_1, \cdots, h_i)$ considered as a function of h_j, such that for the increment η*

$$d_jf(h_1, h_2, \cdots, h_i, \eta) = (g_j(h_1, h_2, \cdots, h_i), \eta).$$

Finally, let $F(h) = f(h_1, h_2, \cdots, h_i)_{h_1 = \dots = h_i = h}$. Then

$$dF(h, \eta) = \sum_{j=1}^{i} d_jf(h_1, h_2, \cdots, h_j, \eta)_{h_1 = h_2 = \dots = h_i = h} \tag{2.11}$$

and

$$\operatorname{grad} F(h) = \sum_{j=1}^{i} \operatorname{grad}_j f(h_1, h_2, \cdots, h_i)_{h_1 = \dots = h_i = h}. \tag{2.12}$$

LEMMA 2.5. *If, in addition to the assumption of Lemma 2.4, $f(h_1, h_2, \cdots, h_i)$ is symmetric in its arguments, then*

$$d_1f(h, h, \cdots, h, \eta) = d_2f(h, h, \cdots, h, \eta) = \cdots = d_if(h, h, \cdots, h, \eta),$$

$$\operatorname{grad}_1 f(h, h, \cdots, h) = \operatorname{grad}_2 f(h, h, \cdots, h) = \cdots = \operatorname{grad}_i f(h, h, \cdots, h),$$

and consequently (see (2.11), (2.12)):

$$dF(h, \eta) = i\, d_if(h, h, \cdots, h, \eta), \quad \operatorname{grad} F(h) = i \operatorname{grad}_i f(h, h, \cdots, h). \tag{2.13}$$

LEMMA 2.6. *Let $f(x)$ possess an i-th differential $d^if(x, h_1, h_2, \cdots, h_i)$ and let $g^i(x, h_1, h_2, \cdots, h_{i-1})$ be the i-th gradient of $f(x)$ (Definition* 2.3).

Then

$$\text{grad}_h\, d^i f(x, h, \cdots, h) = i\, g^i(x, h, \cdots, h), \tag{2.14}$$

where grad_h *means the gradient operation with respect to the variable* h.

Proof. $d^i f(x, h_1, \cdots, h_{i-1}, h_i)$ is linear in h_i and therefore equal to its own differential with respect to h_i and with the increment h_i. Consequently by the definition of grad_i as the gradient with respect to h_i we have

$$\begin{aligned} d^i f(x, h, \cdots, h_{i-1}, h_i) &= d_i d^i f(x, h_1, \cdots, h_{i-1}, h_i) \\ &= (\text{grad}_i\, d^i f(x, h_1, \cdots, h_{i-1}, h_i), h_i). \end{aligned}$$

Comparison with (2.5) shows that

$$g^i(x, h_1, \cdots, h_{i-1}) = \text{grad}_i\, d^i f(x, h_1, \cdots, h_{i-1}, h_i).$$

We now set $h_1 = h_2 = \cdots = h_{i-1} = h$ and apply (2.13) with $f(h_1, \cdots, h_i)$ replaced by $d^i f(x, h_1, \cdots, h_i)$ and obtain immediately (2.14).

Without proof we state the "Leibnitz rule."

LEMMA 2.7. *If the maps* $\gamma(x)$, $\delta(x)$ *of* $C \varepsilon E$ *into* E *possess j-th differentials then the scalat product* $f(x) = (\gamma(x), \delta(x))$ *has a j-th differential and*

$$d^j f(x, h) = \sum_{r=0}^{j} \binom{j}{r} (d^r \gamma(x, h), d^{j-r} \delta(x, h)).$$

Definition 2.6. Let $I = I(x)$ be a real-valued function defined in some neighborhood U of $x = o$ and p an integer $\geqq 2$. Then $I(x)$ is called non-degenerate of order p at o if the differentials of I up to and including order $p + 2$ exist and are continuous in U, if the differentials of order $1, 2, \cdots, p - 1$ are 0 at $x = o$, and if the form $d^p I(o, h)$ in h of order p is non-degenerate in the sense of Definition 2.5. If, in addition, $d^p I(o, h)$ is strictly non-degenerate then I is called strictly non-degenerate of order p.

LEMMA 2.8. *Let* $I(x)$ *possess continuous differentials in some neighborhood* U *of* o. *Then* I *is non-degenerate of order* p *at* o *if and only if*

$$g^1(o, h) = g^2(o, h) = \cdots = g^{p-1}(o, h) = o, \tag{2.16}$$

$$g^p(o, h) \neq o \text{ for } h \neq o. \tag{2.17}$$

Moreover if (2.17) *is replaced by*

$$\| g^p(o, h) \| \geqq \mu \text{ for } \| h \| = 1 \tag{2.18}$$

for some positive μ *we obtain necessary and sufficient conditions for* I *to be strictly non-degenerate at* o *of order* p.

Proof. Suppose (2.16) and (2.17) are satisfied. From (2.16) and the definition (2.5) of the gradient function it follows that

$$(2.19) \qquad d^1I(o,h) = d^2I(o,h) = \cdots = d^{p-1}I(o,h) = 0.$$

Moreover from (2.17) and Lemma 2.6 we see that $\operatorname{grad}_h d^pI(o,h) \neq o$ for $h \neq o$, i. e. (Definition 2.5), that the form $d^pI(o,h)$ is non-degenerate. Thus (2.16) and (2.17) imply that I is non-degenerate of order p. If (2.17) is replaced by (2.18) then Lemma 2.6 shows that $\| \operatorname{grad}_h d^pI(o,h) \| \geqq \mu p$ and we see (Definitions 2.5 and 2.6) that I is strictly non-degenerate at $x = o$.

Conversely, suppose that I is non-degenerate of order p. Then the equations (2.19) hold identically in h, and therefore $\operatorname{grad}_h d^iI(o,h) = o$ for $i = 1, 2, \cdots, p-1$, which, by Lemma 2.6, implies (2.16). Moreover under our present assumption $d^pI(o,h)$ is non-degenerate, i. e., $\operatorname{grad}_h d^pI(o,h) \neq o$ for $h \neq o$ (Definition 2.5), which, again by Lemma 2.6, implies (2.17). In the same way this lemma shows that the strict non-degeneracy of $d^pI(o,h)$ implies the existence of a $\mu > 0$ for which (2.18) is true.

3. Proof of the hypothesis *H* in case of non-degeneracy of order *p*.

THEOREM 3.1. *Let $I(x)$ be strictly non-degenerate of order p at the origin o of the Hilbert space E (Definition 2.6).*

We assume without loss of generality that

$$(3.1) \qquad I(o) = 0.$$

Then the hypothesis H (see introduction) is satisfied.

Proof. We set

$$(3.2) \quad \gamma(x) = \operatorname{grad}\{(x,x)^{p/2}\}; \qquad (3.3) \quad \gamma(x) = px(x,x)^{p/2-1},$$

(3.3) being implied by (3.2).

If $x = h \neq o$ is an element of E such that h and $g^1(h) = \operatorname{grad} I(h)$ are linearly dependent, then (3.3) shows that $\gamma(h)$ and $g^1(h)$ are also linearly dependent. Consequently in the Schwarz inequality

$$|(\gamma(h), g^1(h))| \leqq \| \gamma(h) \| \, \| g^1(h) \|$$

the equality sign will hold:

$$(3.4) \qquad |(\gamma(h), g^1(h))| = \| \gamma(h) \| \, \| g^1(h) \|.$$

In order to prove Theorem 3.1 we will show that (3.4) is impossible for small enough $\| h \| \neq 0$ if

$$(3.5) \qquad I(h) = 0.$$

We will indeed establish the existence of positive constants C, C' such that

$$\| \gamma(x) \| \, \| g^1(h) \| \geqq C \, \| h \|^{2p-2} \tag{3.6}$$

for small enough $\| h \| \neq 0$, while on the other hand for small enough $\| h \| \neq 0$ which in addition satisfy (3.5)

$$| (\gamma(h), g^1(h)) | \leqq C' \, \| h \|^{2p-1} \tag{3.7}$$

Obviously (3.6), (3.7) are (for small enough $h \neq 0$) in contradiction with (3.4), and our theorem will be proved once the existence of constants C, C' with the above properties has been demonstrated.

We start with the proof for the existence of C. In order to estimate the left member of (3.6) we first investigate $g^1(h) = \operatorname{grad} I(h)$. To this end we apply the Taylor formula in the form (2.8), (2.9) with $n = p$ to $f(x) = I(x)$ at $x = o$ and use the equation (2.16) of Lemma 2.7 and (3.1) to obtain

$$I(h) = (g^p(o, h)/p! + R_{p+1}(o, h), \tag{3.8}$$

$$R_{p+1}(o, h) = \int_0^1 (1 - t)^p/p! \; (g^{p+1}(th, h), h) dt. \tag{3.9}$$

To find $g^1(h) = \operatorname{grad} I(h)$ we form the differential $dI(h, \epsilon)$ of $I(h)$ with the increment ϵ. We have from (2.5) and (2.14)[11]

$$\begin{aligned} d_h[(g^p(o, h), h), \epsilon] &= d_h[d^p I(o, h), \epsilon] \\ &= (\operatorname{grad}_h d^p I(o, h), \epsilon) = p(g^p(o, h), \epsilon), \end{aligned}$$

and therefore

$$\operatorname{grad}_h (g^p(o, h), h) = p g^p(o, h). \tag{3.10}$$

Moreover if for any function $I(x, h)$ of x and h, $d_1[I(x, h), \epsilon]$ and $d_2[I(x, h), \epsilon]$ denote the differentials of $I(x, h)$ corresponding to the increment ϵ with respect to x and h respectively we see that $d_h[(g^{p+1}(th, h), h), \epsilon]$ is identical with

$$\begin{aligned} d_h[d^{p+1} I(th, h), \epsilon] &= t d_1[d^{p+1} I(th, h), \epsilon] + d_2[d^{p+1} I(th, h), \epsilon] \\ &= t d^{p+2} I(th, h_1, h_2, \cdots, h_{p+1}, \epsilon)_{h_1 = h_2 = \ldots = h_{p+1} = h} \\ &\qquad + (\operatorname{grad}_h d^{p+1} I(x, h), \epsilon)_{x = th} \\ &= t(g^{p+2}(th, h), \epsilon) + (p + 1(g^{p+1}(th, h), \epsilon), \end{aligned}$$

where again (2.14) has been used. It follows that

$$\operatorname{grad}_h (g^{p+1}(th, h), h) = t g^{p+2}(th, h) + (p + 1) g^{p+1}(th, h), \tag{3.11}$$

---

[11] The index h on d indicates that the differential operation refers to the variable h.

and we obtain finally from (3.8)-(3.11):

$$(3.12)\quad \operatorname{grad} I(h) = g^p(o,h)1/(p-1)! + \int_0^1 (1-t)^p/p![tg^{p+2}(th,h) + (p+1)g^{p+1}(th,h)]dt.$$

Now from Lemma 2.2 follows immediately the existence of a positive constant C_1 such that for small enough $\| h \|$

$$(3.13)\quad \| g^p(o,h) \| \, 1/(p-1)! \leqq C_1 \| h \|^{p-1},$$

while

$$(3.14)\quad \text{norm of the integral in (3.12)} \leqq C_1 \| h \|^p.$$

On the other hand, since I is strictly non-degenerate of order p at o, it follows from Lemma 2.8 (see esp. (2.18)) together with the $(p-1)$-linearity of $g^p(o, h_1, \cdots, h_{p-1})$ that

$$(3.15)\quad \| g^p(o,h) \| \geqq \mu \| h \|^{p-1} \qquad (\mu > 0).$$

Obviously (3.14), (3.15), and (3.12) together imply the existence of a postive constant C_2 such that for small enough $\| h \|$

$$(3.16)\quad \| g^1(h) \| = \| \operatorname{grad} I(h) \| \geqq C_2 \| h \|^{p-1}.$$

This finally proves the validity of (3.6) with $C = C_2 p$ since by (3.3),

$$\| \gamma(h) \| = p \, \| h \|^{p-1}.$$

We turn to the proof of the existence of a $C' > 0$ such that (3.7) holds for small enough h satisfying (3.5). We set

$$(3.17)\quad f(x) = (g^1(x), \gamma(x)),$$

where $\gamma(x)$ is defined in (3.2), and apply the Taylor formula (2.3), (2.4) with $n = 2p-2$ at $x = o$. Since $\gamma(o) = o$ we obtain

$$(3.18)\quad (g^1(h), \gamma(h)) = \sum_{j=1}^{2p-2} d^j f(o,h)/j! + R_{2p-1};$$

$$(3.19)\quad R_{2p-1} = \int_0^1 (1-t)^{2p-2}/(2p-2)! \, d^{2p-1} f(th,h)\,dt.$$

We claim first that all terms of the sum in (3.18) are zero except for the last one, i.e., that

$$(3.20)\quad d^j f(o,h) = 0 \qquad \text{for } j = 1, 2, \cdots, 2p-3.$$

To prove this we apply the Leibnitz rule (2. 15) with $\delta(x)$ replaced by $g^1(x)$. We see from (3. 3) that

$$d^1\gamma(x,h) = ph(x,x)^{(p-2)/2} + p(p-2)x(x,x)^{(p-4)/2}(x,h).$$

From this one proves easily by induction that $d^r\gamma(x,h)$ is a linear combination with constant coefficients (i. e., coefficients independent of x and h) of terms of the form

$$h(x,x)^{\frac{1}{2}\alpha_1}(x,h)^{\alpha_2}(h,h)^{\frac{1}{2}\alpha_3} \text{ and } x(x,x)^{\frac{1}{2}\beta_1}(x,h)^{\beta_2}(h,h)^{\frac{1}{2}\beta_3}, \tag{3. 21a}$$

where α_i, β_i are non-negative integers satisfying

$$1 + \alpha_2 + \alpha_3 = \beta_2 + \beta_3 = r, \quad 1 + \beta_1 + \beta_2 = \alpha_1 + \alpha_2 = p - r - 1. \tag{3. 21b}$$

It follows that $d^r\gamma(o,h) = o$ if $\alpha_1 + \alpha_2 = p - r - 1 > 0$, i. e., for $r = 0, 1, \cdots, p-2$, and (2. 15) (with $\delta = g^1$) gives

$$d^jf(o,h) = \begin{cases} 0 & \text{for } j = 0, 1, \cdots, p-2, \\ \sum_{r=p-1}^{j} \mathcal{B}^j{}_r(d^r\gamma(o,h), d^{j-r}g^1(o,h) & \text{for } j \geqq p-1, \end{cases} \tag{3. 22}$$

where the $\mathcal{B}$'s denote the binomial coefficients. We are interested in j-values $\leqq 2p-3$ (cf. 3. 20); for these j-values and the r appearing in the sum of (3. 22) we have $0 \leqq j - r \leqq 2p - 3 - (p-1) = p - 2$. Therefore the right-hand member of (3.22) will be seen to be zero for $j \leqq 2p-3$, i. e., (3. 20) will be proved, once it is shown that

$$d^sg^1(o,h) = o \qquad \text{for } s = 0, 1, \cdots, p-2. \tag{3. 23}$$

To prove (3. 23) we have only to observe that by Lemma 2. 3(b), $d^sg^1(x,h) = g^{s+1}(x,h)$. But $g^{s+1}(o,h) = o$ for $s = 0, 1, \cdots, p-2$ by Lemma 2. 8 (equ. 2.16), which proves (3. 23).

Thus (3. 20) holds and (3. 18) simplifies to

$$(g^1(h), \gamma(h)) = d^{2p-2}f(o,h)/(2p-2)! + R_{2p-1}. \tag{3. 24}$$

We apply (3. 22) for $j = 2p-2$. Then $j - r = 2p - 2 - r \leqq p - 2$ for $r \geqq p$. This together with (3. 23) shows that (3. 22) reduces to

$$d^{2p-2}f(o,h) = \mathcal{B}^{2p-2}{}_{p-1}(d^{p-1}\gamma(o,h), d^{p-1}g^1(o,h)). \tag{3. 25}$$

Now (3. 21) shows that $d^{p-1}\gamma(o,h) = C'_1h(h,h)^{\frac{1}{2}\alpha_3}$, where $\alpha_3 = p-2$ and C'_1 is a constant. Consequently we obtain from (3. 25)

$$d^{2p-2}f(o,h) = \mathcal{B}^{2p-2}{}_{p-1}C'_1(h,h)^{\frac{1}{2}p-1}(h, d^{p-1}g^1(o,h))$$

and by Lemma (2. 3b)

$$d^{2p-2}f(o,h) = \mathcal{B}^{2p-2}{}_{p-1}C'_1(h,h)^{\frac{1}{2}p-1}(g^p(o,h),h). \tag{3.26}$$

If now (3. 5) is satisfied we see from (3. 8), (3. 9), Lemma 2. 2 and the Schwarz inequality that

$$|(h, g^p(o,h))| = \left| \int_0^1 (1-t)^p (g^{p+1}(th,h),h)\,dt \right| \leqq C'_2 \, \| h \|^{p+1}$$

for some positive constant C'_2. Therefore, (3. 26) shows that

$$| d^{2p-2}f(o,h) | \leqq C'_3 \, \| h \|^{2p-1}$$

for a suitable $C'_3 > 0$. From this, (3. 24), (3. 19) and Lemma 2. 2 follows now obviously (3. 7) for some $C' > 0$.

UNIVERSITY OF MICHIGAN.

REFERENCES.

[1] A. B. Brown, "Relations between the critical points of a real analytic function of n independent variables," *American Journal of Mathematics*, vol. 52 (1930), pp. 251-270.

[2] L. M. Graves, "Riemann integration and Taylor's theorem in general analysis," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 163-177.

[]3 T. H. Hildebrandt and L. M. Graves, "Implicit functions and their differentials in general analysis," *ibid.*, vol. 29 (1927), pp. 127-153.

[4] M. Kerner, "Die Differentiale in der allgemeinen Analysis," *Annals of Mathematics*, vol. 34 (1933), pp. 546-572.

[5] S. Mazur and W. Orlicz, "Grundlegende Eigenschaften der polynomischen Operationen," *Studia Mathematica*, vol. 5 (1934), pp. 50-68.

[6] M. Morse, *The Calculus of Variations in the Large*, American Mathematical Society Colloquium Publications, vol. 18 (1934).

[7] E. H. Rothe, "A relation between the type numbers of a critical point and the index of the corresponding field of gradient vectors," *Mathematische Nachrichten*, vol. 4 (1950/51), pp. 12-27.

ON THE EMBEDDING OF HYPERBOLIC LINE ELEMENTS; A CORRECTION.*

By PHILIP HARTMAN AND AUREL WINTNER.

In our papers appearing in vol. 72 (1950), pp. 553-566 and vol. 73 (1951), pp. 876-884 of this *Journal*, which will be referred to as [1] and [2], respectively, we were dealing with the problem of local embedding of a binary ds^2 into a Euclidean 3-space in the three cases $K > 0$, $K < 0$, $K \equiv 0$ for the Gaussian curvature K of the ds^2. We now see that the treatment of the second of these three cases, that is, of the *hyperbolic* case ($K < 0$), is vitiated by the proof given for the hyperbolic cases of Lemma 1 in [1]. The error is introduced at the end of the last sentence in that proof, lines 11-12 of p. 557, where it is implied that the integral representation of the solution of a hyperbolic differential equation by Riemann's function will produce a certain degree of differentiability. That such cannot be the case follows from the existence of " discontinuity waves " of any given order.

This has no bearing on the treatments of the elliptic ($K > 0$) and parabolic ($K \equiv 0$) cases in [1] and in [2]. Thus, the elliptic and parabolic cases of the Theorem in [1], p. 554, are not affected, nor are those statements in [2] which deal with the elliptic and parabolic cases (namely, (I) on pp. 876-877 and (iii) on p. 882) and the general theorems in [2], namely (i) on p. 879 and (ii) on p. 880.

On the other hand, the hyperbolic case can today be treated only by making use of the general theory of hyperbolic systems, which leads to comparatively high C^n-assumptions, and our method does not contribute anything to this case.

THE JOHNS HOPKINS UNIVERSITY.

* Received November 28, 1951.